

Durham Research Online

Deposited in DRO:

11 February 2008

Version of attached file:

Published Version

Peer-review status of attached file:

Peer-reviewed

Citation for published item:

Castagnoli, L. (2007) 'Everything is true, everything is false : self-refutation arguments from Democritus to Augustine.', *Antiquorum philosophia*, 1 . pp. 11-74.

Further information on publisher's website:

<http://www.libraweb.net/Documenti/Antiquorumphilosophia.pdf>

Publisher's copyright statement:

Additional information:

Use policy

The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a [link](#) is made to the metadata record in DRO
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Please consult the [full DRO policy](#) for further details.

«EVERYTHING IS TRUE», «EVERYTHING IS FALSE»:
SELF-REFUTATION ARGUMENTS
FROM DEMOCRITUS TO AUGUSTINE*

LUCA CASTAGNOLI

*Insignem continent Veritatis astum hae
demonstrationes, quo illa hostium suorum armis in
eorum perniciem pro se abutitur; sed imprimis
Dilemmate velut incantamento cogitur Veritatis
hostis de industria, pro Veritate tamquam pro aris
et focis dimicare.*

ARNOLD GEULINCX

1. PRELIMINARIES

AMONG the theses which ancient philosophers charged with self-refutation (more precisely, with what modern readers have tended to identify with ‘our’¹ self-refutation), we can single out a quite homogeneous class including «Everything is true», «Everything is false», and other similar theses («Every appearance is true», «To say the false is impossible», «Nothing is true», «Truth could perish», etc.). The ancient arguments denouncing such theses as incurring self-refutation will be the protagonists of this article.²

Before starting our inspection some preliminary remarks are needed. It is worth noticing that those theses have attracted very meagre attention in the modern studies on self-refutation: as we shall see, only «Nothing is true» was taken into account by Mackie in his influential formal analysis, only to be neglected in the subsequent section exploring the philosophical dividends of that analysis. The reason for this disinterest is easy to diagnose: such theses are bound to sound to modern ears too blatantly absurd,

* This article is an abridged version of the first part of my Cambridge Ph.D. dissertation *The Logic of Ancient Self-Refutation: From Democritus to Augustine* (CASTAGNOLI 2005), in which I provide a comprehensive analysis of the history and logic of ancient self-refutation.

I wish to express my deepest gratitude to Nick Denyer, for the great skill and generosity with which he supervised my research, and to my examiners, Myles Burnyeat and David Sedley, for their invaluable comments and suggestions and for the encouragement manifested on various occasions. In the last few years I have incurred several other debts of gratitude towards many who, in various ways, contributed to the research of which this article is an offshoot. In particular, I wish to thank Walter Cavini, Valentina Di Lascio, Paolo Fait, Geoffrey Lloyd, Alex Long, Tony Long, Mauro Nasti De Vincentis and Robert Wardy. St. John’s College and Magdalene College, Cambridge, offered me the best conditions for pursuing my doctoral and post-doctoral research, and I am glad to express here my gratitude. A final acknowledgement goes, again, to Myles Burnyeat: although on occasion I shall criticise his views and try to improve on them, the extent to which my work remains indebted to his seminal articles on ancient self-refutation cannot but exceed my numerous acknowledgements *ad locum*, and will not escape the notice of any reader who is already remotely familiar with this topic.

¹ For a disclaimer on the use of «our» here cf. p. 14.

² By «ancient philosophers» and «ancient arguments» I refer here exclusively to ancient *western* philosophers and their arguments. For analogous arguments in the eastern tradition cf. GRAHAM 1989, pp. 183-186 and HARB-SMEIER 1998, pp. 344-345.

and thereby hopelessly uninteresting, to be worthy even of scrupulous refutation. The problem of the nature of falsehood and the puzzle of how it is possible to think and say something false, on the contrary, were live issues for long time in antiquity, and among the priorities on the philosophical agendas of no less thinkers than Plato and Aristotle.¹ This is why radical views such as «Nothing is false» and «Everything is true» were not only proposed by some, but also thought to require and deserve refutation by their opponents. Also the contrary position, according to which nothing is true, found its eager supporters, not only in the now obscure Xenias,² but also in the tradition, ‘sceptical’ and ‘nihilistic’, embodied (at least according to some ancient interpretations) by renowned figures such as Gorgias, Monimus, and Anaxarchus.

From this perspective, we should not confuse «Everything is false» and «Everything is true» with the Liar and the Truth-Teller. The latter were *arguments* (often branded as σοφίσματα, «sophisms», or ἀποροι λόγοι, «insoluble arguments»), perceived by ancient logicians as a serious menace to the foundations of logic, whereas, however surprising this might appear to us, the former were advanced as genuine philosophical *theses*, never making their appearance in any ancient list of sophisms or *insolubilia*. While the reconstruction of the ancient responses to the threat posed by the Liar and the Truth-Teller is extremely controversial, and, given the scantiness of our evidence, deeply conjectural,³ we are sufficiently well informed on the ancient reactions to «Everything is false», «Everything is true», and analogous theses to provide an analysis of them which aspires to be both accurate and instructive on the nature of ancient logic. What we do know for certain about the ancient Liar is that it gave ancient logicians no fewer headaches than its modern heir has afforded to their descendants: Philetas of Cos is reported to have eventually fallen victim to the sleepless nights devoted to it,⁴ and, if without any such dramatic finale, Chrysippus himself certainly spent enormous efforts to solve it, if the catalogue of his writings seems to attest to no less than twelve works in twenty-three books dedicated to the presentation and defence of his own solution and criticism of others.⁵ We shall discover that, on the contrary, the self-refutation charges against «Everything is false» are very straightforward in their logic, and sometimes depicted as almost trivial by their own proponents: if we judge by the tone of our testimonies, whereas the Liar argument was felt as a real challenge, the «Everything is false» thesis was regarded as an embarrassment only for its naïve supporters.⁶

One might protest that, after all, the latter cannot be less paradoxical than (and should be treated along the same lines as) the ‘Epimenides’ (the notorious claim of Epimenides the Cretan that «All the Cretans say the false»),⁷ which is not plainly false, but

¹ For the history of the problem and its solution cf. DENYER 1991.

² Cf. pp. 51 and 61.

³ Cf. e.g. RÜSTOW 1910, CAVINI 1993, MIGNUCCI 1999, CRIVELLI 2004a.

⁴ Cf. ATHENAEUS 9.64, 34-35.

⁵ Cf. D.L. 7.196-197 and BARNES 1996.

⁶ For analogous reasons I shall not deal with «convertible arguments» (ἀντιστρέφοντες λόγοι), i.e. *argument* patterns which can be turned round against their proponent in such a way that both sides have equal force (cf. e.g. the notorious Corax-Tisias dispute). The proponent of a convertible argument can be opposed by an argument with similar structure but opposite conclusion, but he is not the clear-cut loser in the debate, which has *no obvious solution* («convertible arguments» were often classified, like the Liar, among the ἀποροι). For the difference between «convertible arguments» and self-refutation cf. BURNYEAT 1976a, pp. 67-69; for ample discussion of «convertible arguments» cf. NUCHELMANS 1991, pp. 13, 49, 64-75.

⁷ Cf. PAUL. Tit. 1, 12-13: «One of them [sc. the Cretans], a prophet of their own [sc. Epimenides], said: “Cretans, always liars (ψεῦστα), evil beasts, lazy gluttons.” This testimony is true [sic!].»

either false (if some truth has ever been said by a Cretan) or Liar-paradoxical (if all the other Cretan statements are false, or neither Epimenides himself nor any other Cretan did ever say anything else).¹ I shall suggest that the fact that ancient self-refutation arguments appear totally innocent of this complexity does not betray any logical deficiency of their proposers, but comes as invaluable, albeit indirect, evidence that, unlike the ancient reflections on the Liar and many modern self-refutation arguments, they did *not* aim at establishing the truth-value of certain propositions, but served a different purpose. What this purpose was, how the ancients tried to achieve it, and the crucial difference, in aim and structure, between ancient and modern self-refutation will emerge progressively as we proceed.

2. MACKIE ON THE ABSOLUTE SELF-REFUTATION OF «NOTHING IS TRUE»

More than four decades after its first publication in 1964, Mackie's formal analysis of the logic of self-refutation remains the best one on the market, on account both of its undeniable merits and of the scarcity of subsequent attempts.² Since it has also become, through Burnyeat's partial adoption of it, the unchallenged benchmark in most subsequent studies on ancient self-refutation, I suggest we start from the end of our story, and see what Mackie has to teach us on the self-refutation of «Nothing is true».

After introducing «pragmatic self-refutation», with which we shall not be concerned here, Mackie analyses a second type, which he labels «absolute self-refutation», distinguishing two varieties of it, based on two different properties of the main operators involved: «It is true that» (T) has both properties, and thus is involved, in different ways, in both varieties. To begin with, Mackie lists «It is true that» among the *truth-entailing* operators (with «I know that» and «It can be proved that»), i.e. those operators *d*'s «for which if *dp* is true, *p* itself must be true also» (p. 194). On the basis of this law, Mackie constructs the following argument:

- | | | |
|-----|--|---|
| (1) | $(\forall p)(Tp \rightarrow p)$ ³ | T is truth-entailing |
| (2) | $T(\neg(\exists p)Tp) \rightarrow \neg(\exists p)Tp$ | From (1), by substitution |
| (3) | $T(\neg(\exists p)Tp) \rightarrow (\exists p)Tp$ | Existential generalisation |
| (4) | $\neg T(\neg(\exists p)Tp)$ | From (2) and (3), by destructive dilemma $((p \rightarrow q) \wedge (p \rightarrow \neg q)) \rightarrow \neg p$ |

Mackie clarifies what exactly an argument of this form is meant to prove:

With absolute self-refutation of this sort, an item that would be symbolized by $d(\neg(\exists p)dp)$, such as my knowing that I know nothing [or being true that nothing is true], simply cannot occur. Here we can say that each *proposition* of this form is self-refuting. It must be false; given that *d* is truth-entailing, its form guarantees its falsehood. (p. 195)

¹ Cf. e.g. KOYRÉ 1946, PRIOR 1958. We do not know whether the 'Epimenides' was treated together with the 'Eubulides' («I'm saying the false») in the ancient analyses of the Liar. Analogously, «Everything is true» should be either false or Truth-teller-paradoxical.

² Cf. e.g. JOHNSTONE 1964, BONNEY 1966, STROUD 1968, BOYLE 1972, FINNIS 1977, VANDERVEKEN 1980, STACK 1983, CHAMPLIN 1988, JOHNSTONE 1989, WHITE 1989, PAGE 1992, HERRNSTEIN SMITH 1996, JOHANSSON 2003. PASSMORE 1961, the first rigorous analysis of self-refutation, inspired Mackie and is still worth studying. Relevant material can also be found in the vast literature devoted to «Moore's paradox» and to the «pragmatic paradoxes» stemming from it (cf. e.g. MOORE 1942, GRANT 1958, HINTIKKA 1962, SORESENSEN 1988, GREEN-WILLIAMS 2007).

³ I translate, here and hereafter, the Polish notation adopted by Mackie into a more easily readable notation.

Mackie's argument looks unimpeachable. Careful inspection, however, reveals that it is affected by a potentially dangerous ambiguity. It is by no means clear how exactly Mackie wants us to understand the conclusion (4), of which he presents at least two different paraphrases:

- (4^{*}) It is not the case (and it cannot be the case) that it is true that nothing is true;
- (4^{**}) The proposition «It is true that nothing is true» is (necessarily) false.

The former declares the impossibility of a certain state of affairs obtaining, the latter the necessary falsehood of a proposition. (4^{*}) and (4^{**}) are, of course, strictly related: ordinarily we would have no qualms about subscribing to their equivalence (an exemplification of Tarski's T-schema), but we shall discover shortly that their difference can turn out to be significant in certain cases. Mackie's wavering understanding of the main negation is not the only ambiguity to be detected in his double paraphrase of (4). T is taken sometimes as the *sentential operator* «It is true that» (equivalent, I presume, to «It is the case that») and sometimes as the *truth-predicate* «is true» (attached to propositions, sentences, or whatever one might decide the truth-bearers are):¹ «It is true that nothing is true», Mackie's own explicit interpretation of $T(\neg(\exists p)Tp)$, is an odd hybrid of these two understandings of T. Consistency would require either

- (1) It is the case that nothing is the case

or

- (2) (The proposition) «Nothing (i.e. no proposition) is true» is true

and, again, (1) and (2), although strictly related, do not express the same thesis (e.g. (2), unlike (1), is committed to the existence of entities like propositions and truths). One could protest that this indeterminacy is not, after all, very disturbing, since Mackie's conclusion appears to be sound (and soundly inferred) under *all* possible interpretations (provided one interprets the premisses accordingly). It will shortly become clear why this kind of relaxed attitude is not to be recommended.

But let us grant that the argument sketched above is acceptable as it stands: is it a proof that $T(\neg(\exists p)Tp)$ is necessarily false (or «cannot occur») by *self-refutation*? Mackie's argument shows that this formula entails both members of a contradiction: under one of its possible interpretations, for example, the proposition «“Nothing is true” is true» entails both «Nothing is true» (by the truth-entailing property of T) and «Something is true» (by existential generalisation: if «Nothing is true» is true, then certainly there is something true, this very proposition). Is this a *sufficient* condition for self-refutation? Since Mackie does not offer us any *general definition* of self-refutation, and, surprisingly, there is no such definition agreed upon in the literature,² providing a clear-cut answer is a fuzzy task (Socratically, one could despair that it is indeed an impossible task). However, I suggest that such cases are best kept distinct from self-refutations, and best catalogued as *self-contradictions*;³ Mackie himself comes close to admitting as

¹ As we have seen, Mackie's official account treats T, like all the other *d*'s, as an *operator* (p. 193).

² The self-refutation jargon seems to be used (and sometimes abused) with a myriad of different senses and nuances, and those who adopt it rarely make any effort to explain its exact import. Even the studies devoted to the logic of self-refutation (cf. p. 13n2) cautiously eschew the task of providing us with definitions encompassing the various forms they identify and disentangle.

³ The way I intend the latter notion is itself hard to specify in satisfactory formal terms: I take self-contradiction to include all those cases in which a *single proposition*, simple or compound, either entails or consists of a pair

much when he recognises that detecting absolute self-refutations with truth-entailing operators is not philosophically fecund since «we are merely avoiding logical contradictions» (p. 203).

It is time now to consider Mackie's second and more interesting type of absolute self-refutation involving the operator (predicate?) T . T is not only truth-entailing, but also a member of the sub-class of d 's «which we may call *prefixable*, that is ones for which if p itself is true, dp must also be true» (p. 195). On T -prefixability Mackie erects the following proof:

- | | | |
|-----|---|---|
| (1) | $(\forall p)(p \rightarrow Tp)$ | T -prefixability |
| (2) | $\neg(\exists p)Tp \rightarrow T(\neg(\exists p)Tp)$ | From (1), by substitution |
| (3) | $T(\neg(\exists p)Tp) \rightarrow (\exists p)Tp$ | Existential generalisation |
| (4) | $\neg(\exists p)Tp \rightarrow \neg(\neg(\exists p)Tp)$ | From (2) and (3), by transitivity and double negation |
| (5) | $\neg(\neg(\exists p)Tp)$ | From (4), by the logical law $(p \rightarrow \neg p) \rightarrow \neg p$ ¹ |

How should we construe the conclusion (5)? Given the ambiguities I have pointed out above, Mackie seems to be committed to four different interpretations, which he must consider all sound:

- (a) It cannot be the case that nothing is the case;
- (b) It cannot be the case that nothing (i.e. no proposition) is true;
- (c) The proposition «Nothing is the case» is necessarily false;
- (d) The proposition «Nothing (i.e. no proposition) is true» is necessarily false.²

At least two of these immediately strike me as dubious. How could Mackie prove anything like (b)? Suppose, someone as sophisticated as the mediaeval philosopher John Buridan might protest, that God had annihilated all true propositions:³ doubtless no proposition would be true, therefore that nothing is true, although not *possibly-true*, seems to be *possible*, i.e. something which can be the case.⁴ Mackie's conclusion asks us to accept that, somehow, such possibility is *logically* barred: but it is difficult to see why this should be the case (did Mackie establish, as a remarkable by-product of his argument, that necessarily either God does not exist or is not omnipotent?). Reading (d) is no less problematic. Certainly the proposition «No proposition is true» cannot be true,

of contradictory propositions. This broad category would include instances both of *formal* self-contradictions, either explicit (e.g. $p \wedge \neg p$, «It is raining and it is not raining») or implicit (e.g. $(p \rightarrow q) \wedge p \wedge \neg q$, «If it is day, it is light, and it is day, and it is not light»), and of *analytic* self-contradictions (e.g. «This triangle has four sides»). Self-contradictions are also, intuitively, necessary falsehoods (and typically falsifiable through *reductio ad absurdum*), and are treated as such in most logical systems. One might argue that self-refutation must be a subclass of self-contradiction: on some analyses, a proposition refutes itself when it entails its own contradictory, and since anything seems to entail itself, any self-refuting p would always entail the contradiction $p \wedge \neg p$. I shall not assess this view at this stage, but I hope it will become clear later why this classification could be problematic.

I suggest that self-refutation should also be carefully distinguished from *inconsistency*, both semantic and pragmatic, and, as I have mentioned on pp. 12-13, from the semantic paradoxes (cf. CASTAGNOLI 2005).

¹ We shall become very well acquainted with this 'law' of classical logic (a form of the so-called *Consequentia Mirabilis*) in what follows (cf. sect. 6. 1).

² Mackie's own unique paraphrase of (5) is «“There are no truths” is absolutely self-refuting and “There are some truths” is necessarily true» (p. 197), which seems to be equivalent to (d).

³ Adopting modern jargon, for Buridan a *propositio* is «a meaningful sentence token (i.e. a particular utterance or inscription), spoken or written with assertive intent» (HUGHES 1982, p. 5). This need not be Mackie's own conception of «proposition», which unfortunately he fails to clarify, but the use of «statement» (p. 194) and the claim that T -prefixability is a «condition of discourse» (p. 202) might suggest that Mackie's «propositions» are quite concrete linguistic items, not unlike Buridan's *propositiones*.

⁴ Borrowing an important distinction which PRIOR 1969 extracted from Buridan's remarks in *Sophismata* 8.

but suppose, again, that no other proposition existed, or that only false propositions existed (something that Mackie's self-refutation argument cannot exclude): «No proposition is true» would be paradoxical, for the very same reasons which make the Liar assertion «I'm saying the false» paradoxical, and thus Mackie's conclusion that it is necessarily false, by self-refutation, appears too hasty.¹

The conclusion (5) has turned out to be much less palatable and to require much more cautious reflection than we might have thought at first glance; but how was it reached? The key step of Mackie's argument, (2), is an exemplification of T-prefixability: on one of its possible interpretations – the one eventually leading to the two readings of the conclusion just discussed – (2) can be paraphrased as «If nothing is true, then "Nothing is true" is true». Is this kernel of the self-refutation argument acceptable? As an instance of T-prefixability, (2) seems to be perfectly sound, on a par with all the other instances of 'semantic ascent': if snow is white then «Snow is white» is true, if $2 + 2 = 5$ then « $2 + 2 = 5$ » is true, etc. My contention, however, is that T-prefixability should not be light-heartedly assumed *here*. Since T-prefixability takes for granted the existence of truth, to employ it to disprove that «Nothing is true» appears question-begging. «Nothing is true» is obviously inconsistent with the law of T-prefixability, and it comes as no surprise that by assuming the latter Mackie can contradict the former. However, although we lack a proper definition of self-refutation to which we can make appeal, it is not idiosyncratic to suggest, minimally, that a self-refutation argument should show how a certain thesis is refuted (whatever this 'refutation' amounts to) by itself alone or, at most, with the help of other assumptions which have been granted, or would need to be granted, by its proponent *in virtue of his commitment to that thesis*. But no cautious supporter of «Nothing is true» would grant T-prefixability, and hence step (2) in Mackie's argument: by advancing his extraordinary thesis he must be at the same time implicitly asking us to revise many of our basic assumptions about truth, and T-prefixability is no doubt among these.

Assessing a 'revolutionary' thesis against an extraneous and hostile 'conservative' setting produces a refutation that is suspiciously too easy. This is not to deny that such a strategy can be very instructive: the boundary between begging the question by tacitly foisting upon you admissions you would never grant, on the one hand, and *changing the subject* by stubbornly refusing to grant me *anything whatsoever* which is commonly recognised as a defining feature of the subject (making your position unintelligible), on the other, is often quite indeterminate. Perhaps, however, this kind of strategy should not be misleadingly presented as a *self-refutation* argument (and certainly not as an uncontroversial and paradigmatic one). But what seems especially difficult to swallow is Mackie's idea that with absolute self-refutation, unlike pragmatic and operational self-refutation, it is the self-refuting propositional content that falsifies itself, all by itself: not only is a supplementary assumption (T-prefixability) required, but the substantial burden of the refutation is carried by it, and not by the alleged self-refuting proposition.²

Furthermore, we have seen that Mackie's tactic delivers, on this occasion, questionable conclusions. Starting from the next section, we shall begin appreciating the difference, in structure and purpose, between his absolute self-refutation and various ancient self-refutation charges against theses like «Nothing is true» and «Everything is false».

¹ For discussion of the kind of problems raised by Mackie's argument cf. also sect. 7.

² For perplexities on Mackie's accounts of pragmatic and operational self-refutation cf. CASTAGNOLI 2005.

3. SETTING THE STAGE:

Dissoi Logoi 4. 6

Democritus might be the first figure whom our sources credit with deploying a clear self-refutation charge against a thesis belonging to the family we are interested in (Protagoras' «Every *παντασία* is true»). It has been plausibly remarked, however, that our late source, Sextus Empiricus (*M* 7.389-390), employs technical jargon and an argumentative structure which «bespeak a more sophisticated consciousness of logical form than we may suppose was to be found several centuries earlier»¹ and represent a legacy of later (in particular Hellenistic) reflections and developments. For this reason I shall consider Sextus' testimony on Democritus' anti-Protagorean argument together with the other Sextan evidence in section 6, cautiously postponing its scrutiny to a more advanced phase of our inquiry.

With Democritus temporarily sidelined, the earliest argument relevant for us could be one contained in the fourth chapter of the untitled anonymous treatise usually referred to as *Δισσοὶ λόγοι* (*Twofold arguments*) from its opening words.² This sophistic-style collection of arguments for and against various theses was included by Diels in *Die Fragmente der Vorsokratiker* and is standardly dated around 400 BC;³ this dating, however, was more recently (and compellingly) questioned as wholly speculative by Conley and Burnyeat, who argued that, as far as our poor evidence can show, the *Dissoi logoi* could have been written *centuries* after the 404 BC (the likely *terminus a quo*).⁴ While granting this cautionary view on the possibility of dating the *Dissoi logoi* with any confidence and precision, I hope it will not appear too conjectural to assume that this work draws ultimately (if only indirectly) on sources belonging to the sophistic milieu of the late 5th-early 4th century BC, or at least represents a quite successful later attempt to mimic them as faithfully as possible. Even if the author should be much later, no evidence suggests that in the short passage in which we shall be interested he might be contaminating the material he is working on with anachronistic insertions.⁵ Therefore, I shall begin my analysis from the *Dissoi logoi*, with no commitments about its actual date or authorship.

We are in the middle of the fourth chapter, «On truth and falsehood»: the author has just presented some arguments in support of the thesis that the true *λόγος* and the false *λόγος* are the same thing (henceforth, «Identity Thesis», IT), and now is ready to offer a series of arguments for the opposite thesis that «the false *λόγος* and the true *λόγος* are different things, differing by name and also in reality». Here is how the first of these arguments runs:

T1 For if one were to ask (*ἐρωτάσαι*) to those who say that the same *λόγος* is false and true which of the two their own *λόγος* is, if <their reply were> «false», it is clear that <the false *λόγος* and the true *λόγος*> would be two things, while if they were to answer (*ἀποκρίναιτο*)⁶ «true», then this very <*λόγος*> would be also false. (DK90B4, 6)

¹ BURNYEAT 1976a, p. 47.

² The phrase occurs also in the opening sentence of the next three chapters.

³ For extensive discussion about the date of the *Dissoi logoi* cf. ROBINSON (1979, pp. 25-41), according to whom «the Δ. Λ. was written some time around 403-395 (the date accepted by most scholars)».

⁴ Cf. CONLEY 1985, BURNYEAT 1998.

⁵ Of course this diagnosis is largely based on my overall assessment of ancient self-refutation, and cannot be vindicated at this stage.

⁶ As ROBINSON (1979, p. 194) notices, this «disconcerting example of a change from plural to singular is not a *hapax* in the *Dissoi Logoi*».

This argument rephrases IT as «the same λόγος is false and true», and the structure of the refutation makes it clear that this must be in turn understood as «every λόγος is (unqualifiedly) both false and true», equivalent to the conjunction of the contraries «Every λόγος is true» and «Every λόγος is false»:

$$(IT) \quad (\forall p)(Tp \wedge Fp).$$

But the original thesis argued for in the first part of chapter 4 had a quite different shape, at least if we judge by the arguments collected in its support: the true and false λόγος are the same because

- the same λόγος can be true, if the event it describes has taken place, or false, if that event has not taken place (2-3);
- the same λόγος can be true, if uttered by a certain person, and false, if uttered by another (4);
- the same λόγος can be true now, and false tomorrow (5).¹

Burnyeat's general qualm that in the *Dissoi logoi* «many of the arguments for and against do not even manage to contradict each other» fits our case nicely: the argument in T1 seems liable to the charge of *ignoratio elenchi*, since the thesis it attacks (IT) is not the same as the one established by the previous set of arguments (that the same λόγος² can turn out to be true or false depending on different circumstances). To be precise, T1's argument does not manage to contradict any thesis which the previous arguments succeeded in establishing: from the opposite perspective, one could hypothesise that the first set of arguments was aimed at proving, unsuccessfully, the outlandish thesis IT which T1 targets.

Two fundamental features of T1's argument immediately leap to the eye: its *dilemmatic* form and its *dialectical* context. We shall encounter both of them repeatedly: the presence of some kind of dialectical context, in particular, seems to underlie all ancient self-refutation arguments, with only very few possible exceptions (hereafter by «dialectical context» I shall intend, loosely, any dialogical situation in which two opposing parties – either individuals or groups, not necessarily facing each other in flesh and blood – advance and support incompatible views and agree to try to settle their dispute through arguments which are structured in the form of question and answer and respond to some shared rational standards or rules, with the purpose of establishing the truth, or at least the relative merits and plausibility of their clashing positions, and not merely of winning the debate at any cost, unlike the case of 'eristics').³ This should come

¹ For the same pattern cf., e.g., ch. 1 of the *Dissoi logoi*:

(1) Illness is bad for the sick, good for the doctors; victory is good for the winner, bad for the loser; etc.;
 (2) <therefore the same thing (e.g. illness, victory) is good and bad>;
 (3) therefore the good and the bad are the same thing.

The inference from (2) to (3) would sound less problematic to Greek ears than to ours because of the well-known fact that in Greek «the X», where X is a neuter singular adjective, can function not only as an abstract (X-ness) but also as a collective (the Xs, the class of the things which are X), much as in the English phrase «the poor»: τὸ ἀγαθόν can therefore be taken to mean both «goodness» and «the things which are good». For a very different conclusion from premisses of the same kind as (1) cf. Pl. *Resp.* v, 479a-d.

² It is difficult to decide whether «the same λόγος ἥ» is best understood as a single sentence-token (as suggested by the argument of sects. 2-3) or as a different token of the same sentence-type (as required by the argument of sect. 4). The argument of sect. 5 seems to be compatible with both options.

³ Of course such a loose working notion would require to be clarified and narrowed case by case.

as no surprise, since ancient logic never lost its well-known original connection with the concrete practice of dialectic and disputation: nevertheless, I think that some confusion has arisen in the literature from disregard for, or underestimation of, the full import of this datum.

Let us reconstruct the details of T1's dialectical exchange. The proponent of IT is faced by his opponent with a dilemma: does he believe that the λόγος expressing IT is false or that is true? The reasoning underlying the first horn is easy to understand: if the supporter of IT answers that his λόγος is false, then he is conceding that the contradictory of his IT is the case (as long as he endorses the platitude $Fp \rightarrow \neg p$), i.e. that, to borrow the opaque but now familiar jargon of the author of the *Dissoi logoi*, the false λόγος and the true λόγος are two different things. On the other hand, if the proponent of IT grasps the second horn («My λόγος is true»), he is thereby confirming that he takes IT to be the case; but if he accepts that every λόγος is both false and true, he must thereby admit that also the λόγος expressing his own thesis IT is (also) false. Here the argument suddenly stops, with no further comment or clarification: but what has it achieved exactly? Under both the horns of his opponent's dilemma, the proponent of IT has been forced into undesirable admissions: in the first case the straightforward denial of his own thesis, in the second the concession that the λόγος expressing IT is itself (also) false, by self-application. It is not difficult to see why both outcomes are deeply embarrassing, and can be interpreted as amounting to a ruinous dialectical defeat, which anyone should be extremely careful to prevent, by refraining, at the outset, from endorsing IT itself.¹

IT turned out to be a loser in debate; does this mean that what IT expresses – its propositional content, we might say – has been proved not to be the case? I suggest that the answer is no, and – what is more important – that the author of the *Dissoi logoi* shows no interest in establishing the latter, different point. Unlike Mackie's absolute self-refutation, which is supposed to prove the necessary falsehood of certain propositions (or the impossibility of certain states of affairs obtaining), T1's argument 'simply' seems to aim at showing the untenability of IT under dialectical scrutiny. Moreover, the structure of the charge embedded in the second horn is significantly different from that of Mackie's argument: if the λόγος «every λόγος is false (and true)» is true, then every λόγος is false (and true), and therefore the λόγος «every λόγος is false (and true)» must itself be (also) false. This kind of *self-application* is what one would expect to find as a prominent trait of *self-refutation*, and we shall discover that this expectation is met by various ancient arguments. However, we have seen Mackie follow a different (and indeed opposite) route: the fundamental step of his absolute self-refutation is not «If nothing is true, then "Nothing is true" is not true either», but «If nothing is true, then "Nothing is true" is true».²

¹ I use «can» because a full-blown supporter of IT could be prepared to subscribe to the idea that his own λόγος is itself, like every other λόγος, both false and true. Such a hardcore position, however, could still be attacked on dialectical grounds: why has the supporter of IT advanced his thesis, if he believes that it is the case no more than it is not? Why has he answered «true» to the dilemma, when he believes that his λόγος is both true and false, instead of immediately asking to reformulate the question more properly, as a trilemma? (On the basis of the letter of the Greek text, I am assuming that both horns are explicit: «Is your λόγος true or false?», and not «Is your λόγος true (or not)?»).

² ROBINSON'S (1979, pp. 193-194) analogy between IT and the Liar is also ungrounded. Levi's suggestion that T1's argument resembles the περιτροπή of S.E. M 8.389 (cf. T20 in sect. 6. 1) is more to the point, but his assumption that therefore it «also derives from Democritus» (1940, p. 298) appears wholly speculative.

My reconstruction of T1 as a dialectical ‘silencer’ of IT’s proponent could be challenged by observing that the text does not make it explicitly clear that the consequents of the two conditionals describe what must be *granted* by anyone who has subscribed to the corresponding antecedents. This might encourage a different reconstruction of our argument, not as a dialectical dilemma, but as a *proof by cases* of not-IT:

| | | |
|-----|--|--|
| (1) | $T(IT) \vee F(IT)$ | Bivalence |
| (2) | $F(IT) \rightarrow \neg IT$ | Semantic descent ($Fp \rightarrow \neg p$) |
| <3> | $T(IT) \rightarrow IT$ | Semantic descent ($Tp \rightarrow p$) |
| <4> | $IT \rightarrow (T(IT) \wedge F(IT))$ | By substitution (self-reference) |
| <5> | $(T(IT) \wedge F(IT)) \rightarrow F(IT)$ | \wedge -elimination |
| (6) | $T(IT) \rightarrow F(IT)$ | From <3>, <4> and <5>, by transitivity |
| <7> | $T(IT) \rightarrow \neg IT$ | From (6) and <2>, by transitivity |
| <8> | $\neg IT$ | From <7>, (2), and (1), by simple constructive dilemma $((p \rightarrow q) \wedge (r \rightarrow q) \wedge (p \vee r)) \rightarrow q$ |

I recommend that the temptation to read T1 along these lines must be resisted, for at least two reasons. On the one hand, this reconstruction forces us to supplement numerous additional steps of which no trace can be found in the text (in particular, the crucial <7> and <8>); on the other, the two key premisses (2) and (6) patently distort the literal sense of T1. The protases of the two conditional sentences are, respectively, «If <they answer> that their own λόγος is > “false”...» and «If they answer <that their own λόγος is > “true”...», and *not* «If their own λόγος is false...» and «If their own λόγος is true...». This strongly invites us to interpret the apodoses accordingly, since the conditional «If they answer that their own λόγος is “false”, then it is clear that the false λόγος and the true λόγος would be two things», if taken verbatim, is a sheer *non sequitur*: obviously it is not sufficient to *say* that *p* is false for *p* not being the case (the same holds, *mutatis mutandis*, for the second conditional). The most plausible way of making any sense of these conditionals, as charity suggests, is interpreting their apodoses as elliptical, in the way I have proposed above when first reconstructing the argument: «If <they answer that their own λόγος is > “false”, then it is clear that <they are conceding that the false λόγος and the true λόγος> are two things»; «if they answer <that their own λόγος is > “true”, then <they must admit that this very λόγος> is (also) false».

We have just come across another trait which characterises several ancient self-refutation arguments: a tendency to elliptical formulations, in which it is not always immediately transparent whether what is on the table is the truth-values and logical consequences of certain propositions or, rather, the tenability and unavoidable commitments of certain positions in dialectical exchanges.

4. PLATO: DIALECTICAL SELF-REFUTATIONS

Although in the second part of *Dissoi logoi* 4 the position under attack had taken the shape of a conjunction of both the theses we are interested in («Every λόγος is true and every λόγος is false»), T1’s self-refutation charge exploited the self-applicability of the second conjunct only, while the first one remained, as it were, logically inert. I shall ex-

¹ Some verb such as ἀποκρίναιτο is clearly presupposed also by the first conditional.

amine now two Platonic passages in which self-refutation charges are levelled against variants of that first conjunct.

I have emphasised above some crucial differences between Mackie's absolute self-refutation of «Nothing is true» and the *Dissoi logoi* dilemma against IT: no similar comparison will be possible in this section, since in his taxonomy Mackie left no place for «Everything is true», «Nothing is false», and analogous theses. I suppose he did not see them as *self-refuting* at all, since, as we shall discover, their refutation requires one to consider some *external* conflicting proposition which either should be accepted exactly in virtue of the supposed truth of those theses (section 4. 2) or must be eventually granted because of some broader kind of dialectical necessity (section 4. 1).¹

4. 1. *Dionysodorus' downfall*

Let us begin with a passage from the *Euthydemus*. The sophist Dionysodorus has just argued, very deftly, for the surprising thesis that contradicting (ἀντιλέγειν) someone else is impossible,² stunning and silencing his interlocutor Ctesippus (285e-286b). Socrates takes over the conversation:

T2 even though I have heard this particular λόγος [sc. that it is impossible to contradict (οὐκ ἔστιν ἀντιλέγειν)] from many and at many times, I'm always amazed (ἀεὶ θαυμάζω). Protagoras and those like him made considerable use of it, and also some still earlier: but it always seems to me that it's something amazing, and that it overthrows not just the other λόγοι, but itself as well (ἐμοὶ δὲ ἀεὶ θαυμαστός τις δοκεῖ εἶναι καὶ τοὺς τε ἄλλους ἀνατρέπων καὶ αὐτὸς αὐτόν). [...] The λόγος amounts to claiming that it is not possible to say what is false (ψευδῆ λέγειν οὐκ ἔστιν), doesn't it? And when speaking either one says the truth or else doesn't say anything at all? (286c1-8)

Why does Socrates believe that the λόγος that οὐκ ἔστιν ἀντιλέγειν has the amazing peculiarity³ of refuting itself? This is surprising:⁴ after all, it proclaims that *nothing* can be refuted, since it is equivalent to, or at least entails (δύναται), that «it is impossible to say what is false».⁵ Socrates immediately infers that, according to the 'Protagorean'⁶ λόγος, false judging (δοξάζειν), false judgement (δόξα), ignorance (ἀμαθία), and ignorant people (ἀμαθεῖς) will not exist either, obtaining Dionysodorus' eager assent to the whole lot (286d). Socrates protests that Dionysodorus must be speaking only for the sake of argument, but his opponent's reply is dry: «But do refute (ἐλεγεῖν) me, then» (286e1). Socrates complains that there cannot be such a thing as refutation if one accepts, with Dionysodorus, that nobody speaks falsely (286e2-3). The underlying charge is devastating: what Dionysodorus has just *done* (challenging Socrates to refute him) is denounced as inconsistent with what he *says*, since by implying that refutation is possible

¹ I have argued in sect. 2 against Mackie's pretension that in his absolute self-refutation it is a single proposition, all alone, which refutes itself. It is true, however, that the fundamental extra assumption required (T-prefixability) is at least supposed to describe a basic, non-contingent trait of the grammar of a predicate («true») included in the allegedly self-refuting proposition.

² Dionysodorus' argument at 285d-286b need not concern us here. For an interpretation of it and parallel passages cf. DENYER 1991, BURNYEAT 2002.

³ I interpret the first καὶ of 286c4 as epexegetic.

⁴ Pace Rankin, according to whom this is «a semantically self-evident point» (1981, p. 25).

⁵ A thesis previously defended by the two brothers (283e-284c).

⁶ The οὐκ ἔστιν ἀντιλέγειν was better known as an Antisthenic warhorse (cf. e.g. ARIST. *metaph.* Δ 29, 1024b32-34, *top.* I 11, 104b20-21 and p. 23n5). Notice the air of paradox of attributing to Protagoras, the author of Ἀντιλογίαι, the view that ἀντιλέγειν is impossible.

it seems to commit Dionysodorus to the idea that falsehood and contradiction are also possible after all (a form of pragmatic inconsistency). Euthydemus comes to rescue of his companion, helping him to avoid tackling Socrates' criticism by picking up and embracing Socrates' own mocking suggestion that since falsehood is indeed impossible, refutation is also impossible, and thus Dionysodorus *cannot* have challenged Socrates to refute him, despite all appearances (286e4-7):¹ since no one is capable of doing what is not (284c), you cannot order what is not, just as you cannot say what is not. Socrates does not lose his patience, and immediately launches a second attack: if ignorance and error (in action, speech and thought) do not exist, «what in heaven's name do you two come here to teach?» (287a1-b1). The sophists' previous claim to be teachers (cf. e.g. 274a) is charged with blatant inconsistency with the οὐκ ἔστιν ἀντιλέγειν thesis they are now defending, and all its corollaries: teaching certainly involves, among other things, *contradicting* and purging the *false beliefs* of the learner and replacing them with true ones, elevating him from mistake and ignorance to knowledge.² But again the two sophists refuse to tackle Socrates' challenge, accusing him of bringing back into the discussion something said at the very beginning of their exchange (their boasts as teachers) only because he is «unable to deal with what is being said presently» (287b2-5). With this move (a shameless refusal to consider diachronic inconsistency a dialectical sin),³ however, they give Socrates a chance to revive his previous charge: what could the sense of the phrase «unable to deal with what is being said presently» be if not «unable to *refute* the present argument» (287b6-c1)? Dionysodorus' challenge is meaningless, if refutation is impossible. Once again Dionysodorus refuses to answer and wants to restore his preferred role of questioner, but Socrates quickly spots another dangerous inconsistency displayed by this very behaviour: the principle on which Dionysodorus refuses to answer must be that *he* is most skilled in discussion and knows, unlike the ignorant Socrates, when an answer is to be given and when not (287c9-d2). Dionysodorus either does not grasp Socrates' allusion or, more likely, as his reaction betrays («You are just babbling...»), pretends not to grasp it, and goes on with questioning: if only animate beings have sense (νοοῦντα), and phrases are inanimate, as Socrates admits, why has Socrates asked the sense (νοοῖ) of Dionysodorus' phrase «unable to deal with what is being said presently» (287d7-e1)? Dionysodorus' argument is yet another piece of sophistry, based on a blatantly homonymous use of the verb νοεῖν,⁴ but Socrates, instead of denouncing its fallaciousness, once again suddenly brings it to bear against its proponent:

T3 Are you saying that I made a mistake or not? Because if I did not make a mistake you will not refute me, no matter how wise you are, and it is you who are unable to deal with the argument. But if I did make a mistake not even then are you right to claim that it is impossi-

¹ Both Socrates' charge and Euthydemus' defence of his companion seem to overlook the possibility of interpreting Dionysodorus' imperative not as committing him to the existence of contradiction and falsehood *in propria persona*, but only as challenging Socrates to do, if he is capable, that very thing which he insists is possible and which Dionysodorus denies. ² Cf. *Tht.* 161d8-e1.

³ One might object that this refusal appears shameless only because of our failure to take into proper account the real nature of the eristic display of the two brothers, who are not bound to consistency between different episodes or 'rounds' of dialectic. In Socrates' eyes, however, such a rejoinder would be tantamount to confirmation of his initial suspicion that the brothers speak only «for the sake of argument».

⁴ With the meaning of «to mean», in one case, and «to think», in the other one. My English translation is not completely faithful to the Greek original, but at least it manages to reproduce the same kind of homonymy (ὁμωνυμία).

ble to make mistakes. And I'm not talking about things you said last year. So, Dionysodorus and Euthydemus, it looks as if this λόγος has made no progress and still, as in the past, throwing down <the others> falls down itself (καταβαλὼν πίπτειν). (287e4-288a4)

The loop is now complete. Dionysodorus' last sophistry has provided Socrates with confirmation of his initial impression that Dionysodorus' λόγος throws itself down: καταβαλὼν πίπτειν¹ is clearly meant to be equivalent to τοὺς τε ἄλλους ἀνατρέπων καὶ αὐτὸς αὐτόν of T2 (286c4). Both phrases have a nice pictorial force: the verbs καταβάλλω, πίπτω, and ἀνατρέπω were, most probably, all borrowed from wrestling jargon.² This introduces us to another typical feature of ancient self-refutation charges: the large use of metaphors and similes to express and illustrate them.³

Dionysodorus has criticised Socrates for speaking as if phrases had sense: but what can the point of this criticism be for one who takes Dionysodorus' stance on contradiction and falsehood? He is faced with a dilemma: if Socrates did not make any mistake by speaking in that way, then Dionysodorus must admit that his censure has been pointless (a dialectical *error*), and he cannot dismiss Socrates' previous inconsistency charges (he still owes him an answer, and thus *he* appears to be «unable to deal with the argument»); if Socrates did make a mistake, then perhaps Dionysodorus won the penultimate round of the dialectical contest, but the οὐκ ἔστιν ἀντιλέγειν thesis and all its companion theses are automatically admitted by him to be false, and he loses the whole match. Either way, Dionysodorus is defeated.⁴

Socrates' argument looks like a knock-out blow: but why should we classify it as a self-refutation argument, as Socrates' metaphors suggest? Sprague's (1962, 19) solution is unconvincing: to endorse theses which make refutation impossible, and then go on in one's daily business of trying to refute everyone certainly is not a commendable position, but what it can be charged with is 'only' pragmatic inconsistency, and not self-refutation. A different hypothesis which could explain Socrates' self-refutation jargon here follows the guidelines of a solution Sprague herself discards: the argument purporting to prove that it is impossible to contradict (or, perhaps, this very thesis) contradicts (or, more precisely, *is meant* to contradict) the commonsense view that contradiction is possible, and thus commits its proponent to the existence of contradiction, refuting itself.⁵ The difficulty with this proposal⁶ is, trivially, that T3's dilemma, which

¹ Καταβαλὼν could be a polemic allusion to Protagoras' work Ἀλήθεια under its alternative title Οἱ καταβάλλοντες (*The downthrowers*; cf. S.E. M 7.60 and sect. 4. 2), but an early currency of this title is far from certain. On the use of the verb καταβάλλω as a metaphor for refuting someone in the agonistic context of debate and on further ancient references (e.g. EURIP. *Bacc.* 202; PLUT. *Peric.* 8; STOB. *eccl.* II, 23.15) cf. LEE 2005, p. 24n31.

² Cf. also Euthd. 277d1-2; Phdr. 256b; Pol. 583b; HAWTREY 1981, p. 70. Some translators take also ὥσπερ τὸ παλαιὸν at 288a4 as referring to some proverbial expression («as the old saying goes») or to some piece of wrestling jargon («in the old phrase of the wrestling school»), but both proposals seem unwarranted.

³ For an early occurrence of the same wrestling jargon (καταβάλλω, πτώμα, κατάβλημα) for self-refutation cf. Democritus' fr. 125. Notice that Protagoras is credited with a Περὶ πάλης (*On wrestling*, D.L. 9.55). For the ample use of similes in ancient self-refutation cf. CASTAGNOLI 2000 and CASTAGNOLI 2005.

⁴ As Chance notices, Socrates' «refutation is also a perfect illustration of a perennial feature of all comic action: comic inversion. Just as the comic playwright presents, for example, a robber robbed or a mugger mugged, so too Plato has presented the refuters refuted» (1992, p. 108).

⁵ I believe that the proposed line of reasoning underlies this compressed passage in DIOGENES LAERTIUS (3.35): «They also say that Antisthenes, being about to read publicly something that he had composed, invited him [sc. Plato] to be present. And on his inquiring what he was about to read, Antisthenes replied that it was «On the impossibility of contradicting.» «How then,» said Plato, «can you write on this subject?», thus teaching that it incurs reversal (περιτρέπεται). For the meaning of περιτροπή cf. sect. 6.

⁶ Accepted by HAWTREY (1981, p. 108).

appears to be meant by Socrates to back his final self-refutation charge, does not appear to work along these lines. I suggest we can preserve the core of this solution only if we understand the term λόγος as referring here not only to the *thesis* that contradiction is impossible (and its corollaries), but also to all the *arguments* advanced in its support, both to establish it *and* to undermine potential counter-arguments (the whole theory or philosophical outlook, we might gloss). By grasping the second horn of Socrates' dilemma in T₃ and thus confirming his allegation that Socrates was mistaken in speaking as if phrases had sense, Dionysodorus would be trying to defuse Socrates' accusation of inconsistency at 287b6-c1; but by overthrowing (or at least *trying* to overthrow)¹ that accusation, he would be at the same time overthrowing his own position too, by unwittingly conceding that mistake and contradiction do exist after all. One could still feel uncomfortable with the idea that in this way Dionysodorus makes himself liable to the different charge of self-refutation, rather than to a second, renewed charge of inconsistency: strictly speaking, it is not the οὐκ ἔστιν ἀντιλέγειν thesis, taken in isolation or together with some other assumption,² that overthrows itself here, but *Dionysodorus* who contradicts himself by counter-arguing in its defence. However, Dionysodorus' sophistic refutation of Socrates, which clashes with his endorsement of that thesis, is not just any old refutation; it somehow stems from that endorsement for a sort of *dialectical necessity*.³ In a dialectical context like the one depicted in the *Euthydemus* it is necessary that you try to undermine the arguments that your opponent advances against the position you are advocating, unless you prefer to give it up and admit defeat: but if your position is that «it is impossible to contradict» then *any* such attempt (*whatever* its precise content and force) will speak against that position, rather than support it, and will transform you into the best, albeit reluctant, ally of your opponent. Although in T₂ Socrates could not possibly have foreseen to what specific argument of Dionysodorus he would later apply the fatal dilemma of T₃, he could be fully confident, from the very beginning, that the sophist, when challenged, in order to defend his λόγος would *have to* attempt *some* argument against him to which the dilemma could be easily applied.⁴

This makes Plato's idea of «self-refutation» here perhaps looser than we might have expected, and the aim of Socrates' argument certainly different from that of Mackie's absolute self-refutation. Socrates has shown why οὐκ ἔστιν ἀντιλέγειν is bound to be a dialectical loser: every attempt to defend it from attacks would amount to an involuntary admission of its falsehood. That thesis is in fact incompatible with dialectic (just as it is incompatible with teaching), because by denying the possibility of falsehood and

¹ Neither ἀνατρέπω nor καταβάλλω in Socrates' formulations of the self-refutation charge should be intended as 'success verbs'. I suggest they must have a conative nuance: Socrates cannot be saying that the arguments in favour of the οὐκ ἔστιν ἀντιλέγειν thesis do *manage* to overthrow the opposite thesis and the arguments in its defence (and then overturn themselves too), but only that they *purport* to overthrow them (cf. p. 41).

² *Contra* BURNYEAT (2002, p. 41): «if it is true that there is no false judgement, but Socrates thinks it is false, then it is false that there is no false judgement». *Contra* also NARCY (1989, p. 80), FINE (1998, p. 201n2), and KAHN (2000, p. 91), who compare this passage to «the *peritropê* or self-refutation of Protagoras in the *Theaetetus*» (cf. sect. 4. 2).

³ MCCABE (1998, p. 155) notices that «the claim that falsehood is impossible does not *directly* imply its own falsehood [...]»; it needs, instead, a more complex dialectical context to be overthrown», but she fails to clarify what this context is in our passage.

⁴ Socrates' remark «I am not talking of things you said last year» can be interpreted not only 'chronologically' (he is applying his dilemma to the argument Dionysodorus has *just* proposed) but also *logically* (that argument is integral part of the current dialectical round, unlike Socrates' previous reference to the two sophists' boast as teachers).

error it destroys the rationale of debating (proving your own position correct and your adversary's wrong). All you can afford with that thesis is the undialectical behaviour of stating it and then remaining silent. And this too only with the crucial proviso that yours is no ordinary assertion: without such a proviso, the bare statement of it could already be charged with 'overthrowing itself', since normally when you assert that *p* you can be taken to be committing yourself to the idea that *p* is true and *not-p* is thereby false.

Needless to say, this is a most unpalatable result for any philosophical position. But has the proposition itself «It is impossible to contradict» thereby been proved to be false? I suggest that the correct answer is, again, no. The fact that Dionysodorus could not help committing himself to the existence of error and contradiction as soon as he decided to enter the public dialectical arena which is built around those notions does not imply, in point of logic, that error and contradiction do really exist.¹ For, although the sophist's downfall was not the result of any preventable error on *his* part, but reflected an objective indefensibility of his thesis in that setting,² one could protest that that arena hosts worthless games governed by rules which have no correspondence with reality itself: contradiction and falsehood are not possible, although we foolishly behave as if they were (but, then, are those who engage in this sterile dialectical game ignorant and mistaken?). I am not suggesting that Plato believed he could not properly establish the stronger point of the 'absolute falsehood' of Dionysodorus' λόγος, but only that his purpose in the *Euthydemus* section we have just analysed was different and more modest.³

4. 2. Protagoras' «most clever» self-refutation⁴

In the last three decades or so very few Platonic dialogues have attracted the same amount of scholarly interest as the *Theaetetus*; no single passage in the *Theaetetus* has managed to excite the same lively debate as Socrates' so-called «most clever» refutation of Protagoras' «Measure Doctrine» (hereafter, MD) at 171a-c. Since MD, however one interprets it, bears an obvious resemblance to the thesis that «Everything is true» and Socrates' argument is usually treated as a paradigmatic example of ancient self-refutation,⁵ I shall dare to offer my own interpretation of the logic of that argument.

4. 2. 1. The Measure Doctrine and the context of Socrates' argument

Immediately after *Theaetetus*' formulation of his first admissible definition of knowledge, «Knowledge is perception» (hereafter, κρ), Socrates remarks that Protagoras used to say the same, although in a different way:

¹ Just as the mere fact that one *says* that IT is false does not prove that the false λόγος and the true λόγος are different (cf. sect. 3).

² In Socrates' own terms, an *old* defect that no one has ever found a way to overcome.

³ Whether he thought that dialectical indefensibility is a telling sign of something more fundamental about the truth-value of the proposition involved is a question which I cannot address in this article (cf. CASTAGNOLI 2005).

⁴ This section is an abridged and slightly revised version of CASTAGNOLI 2004a, to which I refer for all the details which I could not discuss here, and in particular for analysis of other interpretations.

⁵ Some alternative labels: «recoil argument» (NEWMAN 1982), «turning the table argument» (KETCHUM 1992), and, especially, «peritrope argument» (e.g. LEE 1973, BURNYEAT 1976b, CHAPPELL 2006).

T4 Man is the measure of all things, of those which are, that they are, of those which are not, that they are not. (152a2-4)

Protagoras' MD is paraphrased, in turn, as follows:

T5 As each thing appears (φαίνεται) to me, so it is for me (ἐμοί), as it appears to you, so it is for you (σοί). (152a6-8)

Since, as Socrates goes on to explain, «appears» and «is perceived» amount to the same in the case of perceptual objects (152b12-c2), MD implies that «things are for each person (ἐκαστῷ) such as he perceives them» (152c2-3) and thus that perception «is always of what is, and cannot be false, as befits knowledge» (152c5-6). Therefore, Protagoras' MD provides the sufficient supporting ground for Theaetetus' KP: in a world in which a Protagorean epistemology holds good, perception is always of what is the case (and thus always true) *for the perceiver*. In a case in which, «when the same wind is blowing, one of us feels cold, and the other not» (152b2-3), we must say that the wind «is cold for the one who feels cold, and not for the other» (152b8). This move allows Protagoras to insist that neither perception is false and, at the same time, to avoid accepting the sheer contradiction «The wind is both cold and not cold» (thus preserving, we would say, the principle of non-contradiction – hereafter, PNC). But what on earth does «The wind is cold *for Socrates*» mean? This is explained when Socrates introduces, and progressively unfolds in more and more detail, the «Secret Doctrine» (hereafter, SD) which Protagoras allegedly revealed to his pupils (152c10), meant to provide an ontological setting for the epistemological MD. SD is first introduced as the thesis that «there is nothing which is *just one* thing by itself» (152d2-3), to become very soon the quite different thesis that «nothing is one, either one thing or qualified in one way» (152d6):

T6 all the things of which we incorrectly say that they are (εἶναι) are in the process of coming to be (γίνεσθαι), as the result of movement and change and blending with one another. (152d7-e1)

After a preliminary sketch of an SD-based theory of perception (153d7-154a5), SD appears in the new form «everything is change (κίνησις) and there is nothing else besides change» (156a5). A long description of what the world is like according to this theory follows (156a5-157b1): since each perception (αἴσθησις) and its twinned perceived thing (αἰσθητόν), i.e. quality, are generated together *only* on the occasion of the interaction of their two 'parents' (the perceiver and the perceived object), they are relative to both parents. Each perception is of a perceived object no less than it is of (belongs to) a perceiver; more surprisingly, each perceived quality is *for a perceiver* no less than it is (a quality) of a perceived object. There cannot be any perception to which no perceived quality corresponds (every perception is true: perception is infallible), and there cannot be any unperceived quality (every perceptual quality is perceived: perception is «omni-scient»). Αἰσθητόν and αἴσθησις are like twins, and this should ensure that perception is unerring, «as befits knowledge». ¹ Now we can understand what «The wind is cold *for Socrates*» means. Socrates and the wind generate in their encounter twin offspring, coldness and a perception of cold. Coldness quickly moves towards the wind and qualifies it (relatively to Socrates), so that the wind becomes, *for Socrates*, cold; the percep-

¹ They must be *identical* twins, then.

tion of cold quickly moves towards Socrates and qualifies him (relatively to this wind), so that he becomes a feeling-cold Socrates (in respect to this wind). Is the wind warm or cold in some absolute sense? No, nothing (i.e. no 'parent') is qualified in any way by itself: each thing becomes whatever it is in relation to something.

The import of the qualifiers is not the only aspect of MD in need of clarification. MD is a complex generalisation of the form

$$\forall x \forall y \forall F (\text{if } y \text{ appears } F \text{ to } x, \text{ then } y \text{ is } F \text{ for } x),$$

but whereas the domain of the x s is obvious from the beginning (*men* are measures), the domains of the y s and of the F s appear less definite, and progressively expand from the narrow perceptual objects and qualities we have seen so far to include at the end *all* possible objects of judgement, even in the non-perceptual sphere.¹

At 160e the depiction of KP, MD, SD, and their correlations is complete. Socrates presents a series of objections to MD (161c-164b), but also voices a possible disdainful «apology» on behalf of Protagoras, who protests that those objections have been unfair and based on mere plausibility and verbal traps, hints at how *he* would reply to them, and challenges Socrates to attack and refute what *he actually says*, if he is able to (162d-e; 166a-c).² Socrates can present his objections in a continuous speech, or, if he prefers, use his favoured method of question and answer (δι' ἐρωτήσεων), provided he is fair in his questioning (167d4-e1).³ Subsequently, Socrates persuades a reluctant Theodorus to participate, in place of young Theaetetus, in a more mature examination of Protagoras' doctrine, and in particular of the issue whether it was correct on Socrates' part to have Protagoras concede in his 'apology' (166d-167d) that wisdom does exist, but that the wise are superior to others not on the question of what is true or false (everyone is an infallible measure), but on that of what is better/beneficial or worse/harmful (169d3-8): the wise are those who can *change* the appearances, who, when things appear (and therefore are) bad for someone, can produce a better state by making things appear (and therefore be) good for him (166d6-7). Socrates' plan is to obtain Protagoras' agreement (ἁπολογία) in the quickest and safest possible way, «starting from his own λόγος» (169e8-170a1).

4. 2. 2. Socrates' dilemma

Socrates begins by recalling MD:

- (1) what seems to each one also is for him. (170a3-4)

I have explained the scope MD has reached at this point of the dialogue: «what seems (τὸ δοκοῦν) to each one» is *whatever is judged* by each *man*. Socrates obtains then Theodorus' concession that

- (2) everyone agrees that all men believe in the existence of both wisdom (σοφίαν) and ignorance (ἁμαθίαν). (170b6-7)

¹ For detailed analysis of the various stages of this expansion and discussion of different interpretations of MD, according to which Protagoras is not a relativist «about truth», but «of fact» (e.g. WATERLOW 1977), or is not a relativist at all, but an «infallibilist» or «subjectivist» committed to absolute truths about private perceptual objects (e.g. FINE 1998) cf. CASTAGNOLI 2004a, pp. 5-9.

² For this rejoinder and its intrinsic difficulties cf. *Euthd.* 286e1-3 (p. 21) and sect. 4. 2. 6.

³ On Protagoras' own mastery of both modalities of speech cf. e.g. PL. *Prt.* 329b1-5 and 334e4-335a1.

As evidence for this universal consensus Socrates does not bring the explicit declarations of people, but their behaviour: men look for (or, alternatively, propose themselves as) experts, teachers, and leaders, on the evident assumption that experts, teachers, and leaders are wise in those very spheres in which the laymen are ignorant, and that wisdom is something valuable, sometimes even vital (170a6-b7). Theodorus also grants that

(3) men believe that wisdom is true thinking, ignorance false judgement. (170b9-10)

This ordinary conception of σοφία and ἀμαθία is not the one we have seen Socrates attribute to Protagoras in his ‘apology’: therefore, «men» should be intended only as «most men» (at this stage, fairness requires that Protagoras and his followers are excluded). It is easy to see how (2) and (3) jointly imply that

(4) everyone agrees that (most) men believe that there are false judgements,

from which, as we shall discover shortly, the second premiss of Socrates’ argument is easily secured.

Let us now proceed in our analysis:

T7 SOCR.: What then, Protagoras, are we to make of your λόγος? Are we to say that men always judge what is true, or that they judge sometimes what is true and sometimes what is false? (πρότερον ἀληθῆ φῶμεν εἰ τοὺς ἀνθρώπους δοξάζειν, ἢ τοτὲ μὲν ἀληθῆ, τοτὲ δὲ ψευδῆ;) For, I suppose, from both alternatives it follows that men do not always judge what is true, but both <what is true and what is false>. For think, Theodorus, would anyone of Protagoras’ followers, or you yourself, contend that no one ever thinks that anyone else is ignorant and judges the false?

THEOD.: That’s not a thing one could believe, Socrates.

SOCR.: And yet it is to this that the λόγος saying that man is the measure of all things is necessarily driven.

THEOD.: How is that? (Πῶς δῆ;) (170c2-d3)

Protagoras’ λόγος, which we had found in its usual relativised form only a few lines above in (1), and which we find at the end of T7 in its ‘official’ Protagorean formulation («Man is the measure of all things»), appears different at the beginning of the passage (170c3):

(5) men always judge what is true.¹

The absence of the qualifier «for them» is puzzling, since it seems to make of (5) an «infallibilist» or «subjectivist» thesis, and not that formulation of relativism which we would have expected. Before considering some possible explanations for this absence, let us see what role (5) plays in Socrates’ argument:

- | | |
|---|-----------------------------|
| (a) If (5) men always judge what is true, then men sometimes judge what is false. | $p \rightarrow \neg p$ |
| (b) If men sometimes judge what is false, then men sometimes judge what is false. | $\neg p \rightarrow \neg p$ |
| (c) Either men always judge what is true or men sometimes judge what is false. | $p \vee \neg p$ |
| <hr/> | |
| (d) Therefore, men sometimes judge what is false. | $\neg p$ |

¹ For the reasons why I believe that (5) is meant to be a formulation of MD and discussion of different interpretations cf. CASTAGNOLI 2004a, p. 12.

Whereas both the validity of this constructive dilemma and the truth of premisses (b) and (c) are apparent, the conditional (a), the core of the argument, requires some explanation: that men sometimes judge what is false follows from (5) *and* the further, *external* assumption, already guaranteed by (4), that people believe that there are false judgements. For if every judgement is true, then also the judgement according to which there are false judgements must be true, and therefore there must be false judgements. If one wanted to block this line of reasoning and deny that false judgements exist, one should deny what everyone agrees upon, that people believe in the existence of ignorance and false judgement: but, as T7 reminds us, not even Protagoras or the Protagoreans can be ready to reject this undeniable datum.¹

Socrates' argument, as it stands, is clear, elegant, and unimpeachable: however, *as a refutation of Protagoras* it is vitiated by the fact that the unrelativised (5) does not look like a fair depiction of MD. One might suggest that Socrates is *unwittingly* guilty of *ignoratio elenchi*: surely, however, this is to be ruled out, given that the qualifiers are firm in their place both a few lines above and, as we shall see, a few lines below T7, and it would be impossible for any lucid writer to commit such a mistake, let alone for a Plato.² Might Socrates be dropping the qualifiers *purposely* to get an easy win over Protagoras? One could object, with Burnyeat, that after Protagoras' request to be given a fairer treatment in the discussion of his doctrine, this «would be nothing less than perverse dishonesty», and «perverse dishonesty is not a charge to be leveled lightly against a philosopher of Plato's stature and integrity» (1976b, p. 177).³

How must we interpret T7's argument then? We could understand (5) as an elliptical formulation of

(5*) men always judge what is true *for them*,

supplying the missing qualifier as implicitly meant. This proposal is not as arbitrary as it could appear: there are a few other instances in the *Theaetetus* in which «true», although used within the framework of MD, is not explicitly relativised (161d7, 167a8, 167c2, 172b6), but it is evident that the reader is asked to supply the qualifiers in thought.⁴ If we understand (5) as (5*), Socrates' argument can be reconstructed as follows:

- (a*) If (5*) men always judge what is true for them, then men sometimes judge what is false.
- (b) If men sometimes judge what is false, then men sometimes judge what is false.
- (c*) Either men always judge what is true for them or men sometimes judge what is false.

(d) Therefore, men sometimes judge what is false.⁵

¹ In any case, by *denying* what everyone believes, despite acknowledging that people do believe it, they would at the same time be involuntarily admitting the existence of false beliefs.

² This appears even more clear when we consider that the argument «is presented after eight Stephanus pages worth (160e-168c) of close study of arguments [...] all of which Plato evidently takes to be fallacious precisely because [...] these arguments are careless about qualifiers in various ways» (CHAPPELL 2006, p. 112).

³ I shall argue later, however, that this diagnosis needs to be qualified.

⁴ The unrelativised occurrence at 161d7-8 (αὐτὸς τὰ αὐτοῦ ἕκαστος μόνος δοξάζει, ταῦτα δὲ πάντα ὁρθὰ καὶ ἀληθῆ) is particularly enlightening, since it comes only a few lines after a relativised occurrence, at 161d2-3 (ἐκαστῷ ἀληθὲς ἔσται ὃ αὖ δι' αἰσθήσεως δοξάζῃ) and the two sentences seem to be meant as interchangeable formulations of MD.

⁵ The reader will wonder why I have relativised only «true», and not also «false», which makes this new version of the argument rather asymmetrical. It is reasonable to suppose that Socrates' argument is aimed at refut-

Is this new argument sound? Whereas its validity and the truth of premisses (b) and (c*) are, again, unproblematic,¹ it is hard to imagine any rationale for (a*). The structure of the passage suggests that its consequent (the existence of false judgements) should follow from its antecedent (5*), i.e. MD, and (4), the universally agreed fact that (most) men believe that there are false judgements, something on which Protagoras too must agree. But what can be inferred from these premisses is the relativised conclusion that it is true *for those who judge so* that men sometimes judge what is false, where «those who judge so» should include everyone except Protagoras and his acolytes. Consequently, it becomes mysterious how Socrates' argument is supposed to work as a whole, and how it can be meant to establish the unrelativised conclusion (d).

One could infer that the attempt at reading (5) as implicitly relativised has turned out to be a dead end: we have avoided the Scylla of a sound but irrelevant or question-begging argument only at the high price of falling into the Charybdis of sheer unsoundness. There is a textual clue, however, suggesting that we might be on the right route after all. Our uneasiness is Theodorus' uneasiness; we feel in need for clarification, and Theodorus asks for clarification too («How is that?»). If T7's argument were to be interpreted at face value, without adding the qualifiers, it would be quite easy to see how it works, even for a character, like Theodorus, not particularly philosophically-minded. Since Theodorus is asking for clarification, it seems reasonable to suppose that Socrates will proceed by providing it. This is, at any rate, what happens in the *Theaetetus* in all the six other circumstances in which Socrates' interlocutor demands some explanation by uttering $\pi\omega\varsigma\ \delta\acute{\eta}$; (152d1, 154b10, 164c7, 172c7, 199c12, 201a6);² this, or at least some discernible signal of the fact that one is not going to satisfy one's interlocutor's request, is what the possession of a decent amount of dialectical politeness should guarantee, I suppose, in Socrates' time just as today. Most modern commentators do not seem to agree: they suggest that in what follows we will find a brand new argument (or even two new arguments) against MD.³

4. 2. 3. How to relativise truth and falsehood: the transitional passage

T8 SOCR.: When you have decided something in your mind, and express a judgement about it to me, let's grant that, in accordance with his [*sc.* Protagoras'] $\lambda\acute{o}\gamma\omicron\varsigma$, this thing is true for you; but isn't it possible for the rest of us to decide about your decision? Or do we always de-

ing relativism *tout court*, and thus at establishing the existence of judgements which are false *simpliciter*, and not merely false relatively to the judger (which would mean remaining in a broad relativist framework). Therefore, «false» must not be relativised in (d) and in the consequents of (a*) and (b); but then, since it is logical to suppose that (b) is true simply in virtue of its being a duplicated conditional of the form $p \rightarrow p$, «false» in the antecedent of (b), and consequently «false» in the second disjunct of (c*), must not be relativised either. One might even conjecture that Plato left out *all* the qualifiers in T7 to avoid making this asymmetry distractingly obvious.

¹ (c*) lists only the two alternatives relevant to the present discussion: Protagoras' relativism and the laymen's (and philosophers') ordinary view.

² It is noteworthy that in all these cases the need for clarification does not betray any deficiency of Socrates' interlocutor (Theaetetus in five cases, Theodorus in one), but is caused by the elliptical and obscure way in which Socrates has expressed his thought, alluding to something he is still to explain properly.

³ McDOWELL (1973, pp. 169-170) is one notable exception. SEDLEY's (2004, p. 57) interpretation of the structure of the passage is not liable to my objection, since it locates the beginning of the second argument before Theodorus' request, at 170d1-2.

cide that you judge the truth? Or is it rather the case that on every occasion there are countless people who are in conflict with you, judging the opposite (*ἀντιδοξάζοντες*), and believe that you decide and think the false?

THEOD.: Good heavens, yes, Socrates, countless thousands, as Homer says, who give me all the trouble humanly possible!

SOCR.: What then? Do you want us to say that what you judge on those occasions is true for you (*σαυτῷ*), but false for those countless people (*τοῖς δὲ μυρίοις*)?

THEOD.: It looks as if it is necessary, according to the *λόγος*, at any rate. (170d4-e6)

According to Protagoras' MD (*κατὰ τὸν ἐκείνου λόγον, ἔκ γε τοῦ λόγου*), we should say that since Theodorus judges that *p* then *p* is true *for Theodorus*, and since his opponents judge that *p* is false then *p* is false *for them*. Of course there is nothing excitingly new in this: this is exactly what one should have expected given the way MD has been shaped throughout the dialogue. But since only a few lines before, in T7, the Protagorean *λόγος* had apparently altered into a different, non-relativistic thesis, this transitional passage turns out to be important: it supports my hypothesis that the qualifiers should be supplied also a few lines before, in Socrates' dilemmatic argument, unless one prefers to attribute to Plato a quite schizophrenic attitude towards MD (and much more so, if one accepts my contention that T8 is supposed to begin an explanation of how Socrates' dilemma in T7 works).

Having proposed a model treatment of qualifiers in case of conflicting judgements, Socrates applies it to Protagoras' case. He starts from what he takes to be a datum: the vast majority of people do not believe that man is the measure. If Protagoras himself does not believe¹ his own doctrine either, then no one believes it,² and thus, according to MD itself, it is true *for nobody* (170e7-171a1).³ This might look *prima facie* like a counterfactual hypothesis: Socrates will aim at showing that it is not as notional as it could appear. If we suppose, on the contrary, that at least Protagoras does believe his MD, a first consequence is that MD is still false for many more people than it is true for: just as in the previous example about Theodorus, we might say that MD is true only *for Protagoras* but false *for countless thousands*, and thus, *in this sense*, more false than true (171a1-5).

4. 2. 4. The MD's most clever feature

There is a more unexpected consequence coming next:

T9 SOCR.: Secondly, it has this most clever feature (*τοῦτ' ἔχει κομψότατον*): on the one hand, he [sc. Protagoras] concedes (*συγχωρεῖ*), in some way, that regarding his own opinion the opinion of those who judge the opposite (*τῶν ἀντιδοξάζόντων*) (by which they think that he says the false) is true, since he agrees (*ὁμολογῶν*) that all men judge what is the case (*τὰ ὄντα δοξάζειν ἅπαντας*).

THEOD.: Undoubtedly.

SOCR.: And then, if he admits (*ὁμολογεῖ*) the truth of the opinion of those who think that he says the false (*αὐτὸν ψεύδεσθαι*), he is conceding (*συγχωροῦ*) the falsehood of his own opinion?

¹ More precisely, «*did not believe*»: at the dramatic date of the dialogue Protagoras is dead (cf. sect. 4. 2. 6).

² This sounds quite surprising: one would think there must be other Protagoreans around endorsing, or at least *pretending* to endorse, MD. Perhaps one could intend «Protagoras» as «Protagoras and his faction». However, this slight inaccuracy is not too damaging: the same argument which Socrates will use to show that Protagoras can be forced to join the anti-MD consensus could be used against other Protagoreans.

³ «The truth that he wrote» is a pun, referring to MD but also to the work, *Truth* (*Ἀλήθεια*), beginning with it. It also echoes Protagoras' own words at 166c9-d1: «For I do say that the truth is as I have written it».

THEOD.: Yes, necessarily.

SOCR.: On the other hand, the others don't concede (οὐ συγχωροῦσιν) that they say something false (ἐαυτοὺς ψεύδεσθαι)?¹

THEOD.: No indeed.

SOCR.: But he, again, admits (ὁμολογεῖ) that also this belief is true, according to what he wrote?

THEOD.: So it appears.

SOCR.: It will be disputed (ἀμφισβητήσεται), therefore, by everyone, beginning with Protagoras—or rather, it will be admitted (ὁμολογήσεται) by him, when he concedes (συγχωρῇ) to the person who contradicts him that he believes the truth—when he does that, even Protagoras himself will be conceding (συγχωρήσεται) that neither a dog nor just any human being is the measure of anything at all which he hasn't learnt. Isn't that so?²

THEOD.: It is. (171a6-c4)

We have finally arrived at Socrates' «most clever» self-refutation argument. I have adopted this label so far noncommittally, simply because it or some equivalent label suggesting that the argument is particularly «ingenious», «exquisite», or «subtle» are used by most commentators,³ with very few exceptions. However, I abandon this established use here, since it is based, I maintain, on a misreading of 171a6: the unstated subject in ἔπειτά γε τοῦτ' ἔχει κομψότατον must be «the ἀλήθεια that he [sc. Protagoras] wrote» (170e9-171a1), i.e. *MD itself*. It is not Socrates' forthcoming self-refutation argument *against* MD that is singled out as «most clever» at the beginning of T9; it is MD that «has this most clever feature»,⁴ where «this» refers forward to the surprising fact that also its inventor, Protagoras, can be forced into rejecting it through a self-refutation argument. Is being necessarily repudiated by its own creator a sign of particular cleverness for a philosophical thesis? Since the answer must be a definite no, we should understand Socrates' remark as mocking:⁵ after all, also the two other occurrences of the adjective κομψός in the *Theaetetus* (156a3,⁶ 202d10)⁷ do not seem to express Socrates' unmixed praise.

Having bracketed the prejudice that T9's argument *must* be particularly clever, let us examine it. Its most puzzling feature is, again, the absence of qualifiers, beginning with the unrelativised formulation of MD as «all men judge what is the case» at 171a9: we have appreciated how crucial the qualifiers are in Protagoras' λόγος and we have found them in the statement of it introducing the whole section (170a3-4) and again, repeatedly, only a few lines before T9 (170d5, 170e4-5, 170e9). One possibility, powerfully advocated by

¹ I shall argue on pp. 34-35 that this line must be read and interpreted differently.

² For a different construal of this convoluted sentence based on different punctuation cf. SEDLEY 2004, p. 60.

³ E.g. «especially clever refutation» (FINE 1998, pp. 208-209), «the “clever” second phase [of the argument]» (McCABE 1994, p. 278), «very subtle argument» (BOSTOCK 1988, p. 89), «the subtlest argument» (POLANSKY 1992, p. 130), «most subtle argument» (SEDLEY 2004, p. 60), «really exquisite argument» (BURNYEAT 1976b, p. 177), «most ingenious argument» (WATERLOW 1977, p. 19), «argument [...] ingenious» (CHAPPELL 2006, p. 135), «exquisite argument» (e.g. LEE 1973, p. 243; LONG 2004, p. 35; WEDIN 2005, p. 171).

⁴ For the same construction cf. *Phdr.* 275d4, *Chrm.* 172b1. GOTTLIEB (1992, p. 189) seems to construe the text similarly.

⁵ On this interpretation Socrates' remark echoes his claim in the *Euthydemus* that the (Protagorean) λόγος according to which it is impossible to contradict appears always θαυμαστός («amazing») to him, because it «overthrows not just the other λόγοι, but itself as well» (286c3-4; cf. T2 on p. 21). Cf. also *Euthd.* 303d-e.

⁶ The proponents of SD are «much more subtle» (πολύ κομψότεροι) than those uninitiates who do not admit the existence of anything except what they can touch and see.

⁷ What seems the «most subtle point» (κομψότατα) in 'Socrates' dream' (201d-202d) is that the elements are unknowable but the complexes knowable.

Burnyeat, is that Plato wanted his readers to realise by themselves that they should restore the missing qualifiers and to understand by themselves how the argument would then work. In sect. 4.2.2 I provisionally adopted the same conjecture as the most plausible in interpreting T7's dilemmatic argument; however, I had to admit that it is mysterious for me (just as for Theodorus) how that argument can remain sound once we have supplied the qualifiers. *Prima facie*, adding the qualifiers does not help us with T9 either: since Protagoras believes that (1) «all men judge what is the case <for them>», he must admit that (2) also the judgement of those who believe that MD is false is true <for them>, and thus that (3) MD is false <for them>. One might argue that (3) is already a dangerous admission for Protagoras to grant, for although it does not imply that MD is absolutely false, it does imply that MD is not absolutely true. This is not surprising: MD cannot be an absolute truth if it falls into its own scope (something which, interestingly, is never called into question in the *Theaetetus*)¹ and thus each man must be the measure also of his own being a measure and of other men's being measures. More surprising is the fact that in the *Theaetetus* Protagoras does seem to consistently assert his MD as an absolute, unqualified truth:² «Man is the measure of all things» (152a2-3), «I do say that the truth is as I have written it» (166c9-d1), «You have to put up with being a measure, whether you like it or not» (167d3-4).³ Protagoras' inescapable admission of (3) is sufficient to compel him to abandon this blunt and inappropriate way of presenting his theory; nonetheless, by granting that MD is false for his opponents, albeit true for him,⁴ Protagoras does not seem committed yet to admitting *in propria persona* that MD is false *simpliciter*, since the idea that MD must be true for everyone, or true absolutely, does not seem to be part of MD itself.⁵

¹ Contra SEDLEY 2004, p. 48.

² Thus violating the rule expressed at 160b8-c1. Analogously, SD is presented as an unqualifiedly true ontology (cf. SEDLEY 2004, p. 48), but this clashes with the relativism such an ontology is supposed to back and justify.

³ Protagoras' book was entitled *Truth*, without qualifications.

⁴ Several interpreters stress the point that this concession is already sufficient to make Protagoras' position utterly uninteresting and solipsistic, and that this is the strongest lesson of Plato's refutation of Protagoras, even if Socrates' argument fails to establish that Protagoras himself must admit the absolute falsehood of MD *in propria persona* (cf. e.g. LEE 1973; WATERFIELD 1973, p. 176; McDOWELL 1973, p. 71; WATERLOW 1977, pp. 35-36; BOSTOCK 1988, p. 95; MCCABE 1994, p. 279; CHAPPELL 2005, p. 114).

⁵ BURNYEAT'S (1976b) contention is that this impression is misguided: once we have understood the import of the admission that MD is false for those who judge it false, we realise that Protagoras cannot concede this and at the same time refuse to grant that MD is false *simpliciter*. I cannot present and discuss Burnyeat's influential proposal here, for which I refer to CASTAGNOLI 2004a, pp. 15-18 (for recent criticism of Burnyeat's reading along similar lines cf. now WEDIN 2005, pp. 174-178). One of my perplexities concerned Burnyeat's use of the metaphor of «private worlds» (something is true for *x* if and only if it is true in *x*'s world), but I feel now that I failed to emphasise the real crux of that use. By excluding the possibility that a world might be incorporated into another as «incoherent», solipsistic, and almost nonsense (1976b, p. 191), Burnyeat is denying the possibility to say that it is true for Protagoras that what Socrates believes is true *for Socrates*. In other terms, denying the possibility of private worlds embedded into other private worlds is the same as denying the possibility of «repeatable qualifiers» (the idea that Protagorean qualifiers must be unrepeatable is argued for by DENYER 1991, pp. 90-94). This «single-relativisation assumption» («no truth is or could be hierarchically relativised to two or more subjects») is used by SEDLEY (2004, pp. 57-62) to vindicate the soundness of Socrates' argument: «When Protagoras is forced to agree that his opponents' view is correct, the reason why this is not qualified as "correct for them" is that his responses are establishing what is true in his own world» (2004, p. 61), since T9's argument is implicitly governed by «What is the case for Protagoras himself?» at 170e7 and double-relativisations are barred. (Sedley adopts the same strategy to explain the apparent lack of qualifiers in T7: he reads τῷ Πρωταγόρῃ at 170c2 with MSS B and D, and takes the whole following dilemmatic argument to be governed by this qualifier [«How then shall we run the argument for Protagoras?»], which bars the addition of further qualifiers.) I am not sure that in the absence of any explicit evidence for this single-relativisation assumption in the *Theaetetus* we are entitled to supply it and to make it the core of Protagoras' refutation

The first step towards the solution to our puzzle will consist in reinterpreting the central lines of Socrates' argument in T9 (171b4-9: «On the other hand [...] so it appears»): their exact function has been left unexplained by commentators until Emilsson, in a thought-provoking article published in 1994, first highlighted their crucial role. Since these lines come immediately before the conclusion, which is introduced by ἔρα («therefore»), one would expect them to make some major contribution towards it. Emilsson notices that it would be unnatural to describe Protagoras' opponents' spontaneous assertion that they are not wrong in maintaining that MD is false as the refusal to *concede* (οὐ συγχωροῦσιν) something (lines b4-5): «in a dialogue context, as here, “don't admit” indeed suggests that Protagoras is supposed to have said something to which the opponents refuse to give their assent», something which Plato fails to record (1994, p. 140). But what might Protagoras have said? Certainly not that his opponents are mistaken (this would amount to a disastrous admission that false judgements do exist);¹ what Protagoras is allowed to retort is that his opponents' view on MD is false *for him* (i.e. that MD is true for him). But his opponents do not concede this, do not concede that they are saying something which, although true for them, is also false for Protagoras: since they do not believe MD, the very concepts of relative truth and falsehood are unacceptable for them. And Protagoras himself cannot help admitting (ὁμολογεῖ), in accordance with his own theory, that also this judgement is true (lines b7-8). True for whom? Surely only for them? No, since according to Emilsson Protagoras has now been 'disarmed' of his qualifiers: «the objection that Plato's argument depends on ignoring the qualifiers is misplaced», since the function of lines b4-5 «is precisely to show that Protagoras' opponents will not accept the answer which the critics have thought available to him» (1994, pp. 142-143).²

I shall try to improve on Emilsson's idea by presenting a more straightforward reading of lines b4-8 and by elucidating how *exactly* they are meant disarm Protagoras of the qualifiers. My interpretation relies on the adoption, at line b4, of the *lectio* ἐαυτοῖς (MSS B, D, and T) in lieu of ἐαυτούς (MS W), which we find printed in the most recent edition of the *Theaetetus*³ and was accepted by Emilsson. This apparently minor choice

on the basis of an argument *ex silentio* (no double-relativisation occurs in the *Theaetetus*). On such an assumption, any conflict of judgments of which one is aware would commit MD to blatant contradiction: if Protagoras judges that Socrates is pale and also that Socrates believes of himself that he is tanned, then Socrates must be at the same time pale and not pale (tanned) *for Protagoras* (i.e. in Protagoras' world), since Protagoras cannot say that it is true for him that it is true for Socrates that Socrates is tanned. Moreover, one might wonder why Plato should have failed to emphasise the crucial point that double relativisations would be senseless. (More specifically, Sedley's proposal to read αὐτῷ Πρωταγόρᾳ at 170e7 as a Protagorean relativiser appears difficult, especially because Sedley's interpretation requires us to believe that this relativiser does not govern the immediately following lines 170e7-171a1, where other qualifiers occur, begins governing the arguments from 171a1, including T9, and ceases governing the argument again at 171c5 [T10 on p. 37], all of this without any signal in the text.) For recent and insightful discussion and criticism of the single-relativisation assumption cf. also CHAPPELL 2006, pp. 113-120. Chappel's own positive proposal is that Socrates is warranted in dropping the qualifiers because Protagoras' aim is to *reduce* truth to relative truth: «the supposition that the properties of the *analysandum*—truth—transfer across to the *analysans*—truth—leads Protagoras into the contradictory position of accepting that his philosophical opponents' views may just be described not just as true for them, but as true *simpliciter*» (2006, p. 130). However, to refute Protagoras by saddling him with such an extreme brand of reductionism seems to me no less question-begging than treating him as a subjectivist while his position is a relativist one.

¹ Cf. 161e2-3: «we who are ourselves each the measure of his [sc. Protagoras'] own wisdom».

² EMILSSON 1994, pp. 142-143.

³ Hicken's edition in DUKE-HICKEN-NICOLL 1995 (cf. also CAMPBELL 1861, McDOWELL 1973, NARCY 1995). WOHLRAB 1891, BURNET 1900, and DIÈS 1924 read ἐαυτοῖς (but Diès does not translate it). Among the interpreters, only POLANSKY (1992) and BEMELMANS (2002) adopt ἐαυτοῖς and read it as a Protagorean qualifier.

will allow us to reinterpret this portion of the exchange between Protagoras and his opponents in a way resembling Emilsson's without any need for supplying implicit steps in the argument (notably, nothing in Plato's Greek corresponds to Protagoras' retort that his opponents' view on MD is false for him).

By admitting that the opinion of his opponents about MD is true (for them), Protagoras is conceding that MD is false: *false for them*, of course (I am supplying here the missing qualifiers at 171a6-b3 like Burnyeat and Emilsson). But his opponents are not content with this concession: they are not ready to grant this qualification, that he says something which is *false for them* (οὐ συγχωροῦσιν ἑαυτοῖς ψεύδεσθαι). My reading takes αὐτόν (Protagoras) as the unstated subject of ψεύδεσθαι (cf., similarly, αὐτόν ψεύδεσθαι immediately before, at line b2): since, «if there is no specific subject of the infinitive then the indefinite accusative idea takes over [...] but such an indefinite or generic turn of thought is often used when there is in fact a specific reference within the context»,¹ the most accurate translation would be «the others do not concede that one says something false for them»,² where «one» alludes to Protagoras, referred to immediately before. Alternatively, it is not wild speculation that Plato might have originally written ἑαυτοῖς αὐτόν ψεύδεσθαι but αὐτόν was inadvertently dropped out of our manuscript tradition at some early stage. The alternative *lectio* ἑαυτούς is suspect also on purely linguistic grounds: «if the subject of the infinitive is the same as the subject of the leading verb, then, the proper accusative subject is usually displaced by the nominative of the original expression of the idea».³

But why are Protagoras' opponents unwilling to accept Protagoras' qualification that it is for them that MD is false? What they believe is that Protagoras, by advancing his MD, is saying something false *simpliciter*, and not only for them (or indeed for anyone else).⁴ On my interpretation at lines b4-5 the qualifier (ἑαυτοῖς) finally makes its appearance in the text: unlike Emilsson, who makes qualifiers a major issue in T9 in spite of their complete absence, I ask the reader to intend them as implicitly meant only in its first part (at 171a8, a9, b1, b2), where it is not so difficult to accept the integration given that they are in their place both immediately above (170e9: μηδενί) and, on my reading of the text, below (171b4: ἑαυτοῖς).

What can Protagoras reply to his opponents' refusal to qualify their denial of MD? He cannot protest that they are mistaken: according to MD, he cannot help saying that their

¹ COOPER 1998, p. 774.

² Although ψεύδεσθαι in connection with Protagorean qualifiers («to say something false for...») is not used anywhere else in the *Theaetetus*, we have encountered two sufficiently close parallels at 170e4-5 (σαυτῷ ἀληθῆ δοξάζεις and τοῖς μυροῖς ψευδῆ <δοξάζεις>).

³ COOPER 1998, p. 771. Unfortunately, however, this cannot settle the question, since there are exceptions to the general rule, some of which can be found in Plato.

As we have seen, Emilsson, who reads ἑαυτούς at line b4, must supply before it Protagoras' *implicit* reply «But you must admit that your view is false *for me*». Emilsson examines the reading ἑαυτοῖς, but rejects it on the grounds that «not even in Protagorean language does there seem to be anything describable as “being wrong for oneself”» (1994, p. 139n8). This is correct, but Emilsson fails to see the possibility, which I defend here, of taking αὐτόν (Protagoras), and not ἑαυτούς (Protagoras' opponents), as the subject of ψεύδεσθαι.

Bemelmans, who reads ἑαυτοῖς, translates lines b4-5 as follows: «But the others don't concede that it is (true) for themselves, that what he thinks is false» (2002, p. 80). This is consistent with my reading, but I find the addition of «it is true» unnecessary. Polansky's paraphrase is very similar to Bemelmans' translation (1992, p. 131), but he adds that «Socrates' withholding of the qualifying labels “to him” and “to them” is playfully unfair».

⁴ Textual support for this reading might come from 179b7-9, at which Theodorus, referring back to T9, says that Protagoras' λόγος «is refuted also when it makes other people's judgements authoritative, but they clearly think that his theories are *in no way* (οὐδαμῇ) true».

belief that MD is not false for them, but simply false, is true (171b7-8). We would have expected a qualified admission that this second-order belief too is true *for them only*; surely we must supply again the missing relativiser? If we did so we would end up with our original problem unsolved: how could Socrates shift in the lines which immediately follow (b10-c3) from Protagoras' admission that it is true for his opponents that MD is false (and not only for them) to the devastating admission that MD is unqualifiedly false? Emilsson's insight must be correct: at lines b4-8 Protagoras is being disarmed of the qualifiers; the problem is to understand *how*. Lines b4-5 suggest that Protagoras' opponents are not ready to accept *any* relativisation of the truth of their judgements; however, *prima facie* this does not seem to require that Protagoras himself gets rid of the qualifiers. For, apparently, he might continue relativising the truth and falsehood of his opponents' claims, without contesting their refusals to accept such relativisations: in this way Protagoras would be giving up any hope to persuade his opponents, but at least would steal a draw.¹ Emilsson suggests that, since Protagoras never objects to his opponents' refusals to relativise the truth-value of their claims, his opponents «could simply say "Since you have no objection at all, Protagoras, we understand that you have given your admission to our statement"» (1994, p. 145), i.e. *pretend* that Protagoras *himself* has admitted the unqualified falsehood of MD. However, this would be a rather unfair treatment of Protagoras' dialectical behaviour: as long as *he* continues relativising the truth of his opponents' claims (something from which he has not yet been proved to be barred), he should not be charged with endorsing them on the mere grounds that he does not charge them with absolute falsehood (to require him to do this would be only another way of begging the question against MD).²

I suggest that, despite appearances, after his opponents' refusal to accept relativisations described at lines b4-5 Protagoras *cannot* continue adding his qualifiers. For let us reconstruct in direct form how the entire dialectical exchange between Protagoras and his opponents would unfold if Protagoras clinged to his qualifiers:

OPPONENTS: By advancing MD you say something false.

PROTAGORAS: I concede that this view of yours is true for you, since all men believe what is the case for them.

OPPONENTS: So your MD is false.

PROTAGORAS: I concede that it's false for you.

OPPONENTS: But we don't concede *this* to you. We don't believe that MD is false *for us*: it's false *simpliciter*.

PROTAGORAS: I admit that also this belief of yours is true, *for you*. It's true for you that MD is not false for you, but false *simpliciter*.

Protagoras' final remark is only apparently one more harmless admission of the relative truth of his opponents' claim. Since «It is true for you that MD is false» seems to amount to the same as «MD is false for you», Protagoras' attempt to remain faithful to his MD by conceding, once again, the relative truth of his opponents' claim would turn

¹ But in this way he would fail to qualify for wisdom in the sense of the superior skill to change others' appearances (cf. p. 27). According to Bemelmans, continuing qualifying the truth of his opponents' claims is not a viable option because «this would lead to an infinite regress» (2002, p. 82), but he fails to explain why this regress should be dangerous for Protagoras, or at any rate attributable to Protagoras more than to his opponents (they stubbornly continue denying Protagoras' qualifications just as Protagoras continues qualifying their denials).

² WATERLOW (1977, p. 31) is liable to a similar objection.

out to be a straightforward and *unqualified denial* of that claim, oddly presented as an agreement («I admit...»):

PROTAGORAS: I concede that MD is false for you.

OPPONENTS: But we don't concede this to you. We don't believe that MD is false *for us*: it's false.

PROTAGORAS: I admit this too: MD is false, *for you*.

On this occasion relativising would not be, ultimately, a way of granting a qualified acceptance to someone else's belief, without endorsing it *in propria persona*, but a way of openly *contradicting* it, thus denying MD itself.¹ Moreover, Protagoras would be repeating, in slightly different terms, exactly the same thing which he has already said just a few seconds before and which he knows has not (and will not) be accepted by his opponents. In other words, he would be merely 'babbling'.²

It is for these reasons that after lines b4-8 Protagoras must give up his qualifiers, willy nilly, and admit that the belief of his opponents, according to which MD is false, is true.³ This admission is taken by Socrates and Theodorus as evidence that everyone, including Protagoras, disputes MD (171b10-c4): the anti-MD consensus is now universal. What at 170e7-171a1 had appeared as a merely counterfactual possibility («if not even he himself thought that man is the measure...») has turned out to be a necessary dialectical outcome of the clash between Protagoras and his opponents. Even Protagoras will be compelled to reject MD when faced with disagreement. It is crucial to emphasise that the result of the dialectical manoeuvre described in T9 is not (and cannot be) the demonstration of the necessary ('logical') falsehood of MD, but *Protagoras'* admission of MD's falsehood and consequent defeat by 'reversal'.⁴ Grasping this point is fundamental to understanding the rationale of the next and final step of Socrates' argument:

T10 Then, since it is disputed (*ἀμφισβητεῖται*) by everyone, Protagoras' truth is not true for anyone, neither for anyone else, nor for himself. (171c5-7)

MD is still assumed here, along with T9's conclusion that everyone disputes MD, to draw the further conclusion that MD is not true for anyone (Protagoras included): if his previous argument had been recognised by Socrates as a sufficient proof of the MD's absolute falsehood, this final step would be unwarranted or at least redundant, and even more so given that the conclusion that MD is false for everyone sounds like a weaker (and perhaps incompatible) one.

¹ As McCabe notices, «disagreement with everyone else is inaccessible to him [sc. Protagoras]; all he can ever do is agree» (2000, p. 43); but in this case even qualified agreement would be inaccessible, since it would amount to sheer disagreement, dooming Protagoras to inconsistency.

² I adopt this term in the narrow sense in which Aristotle uses it at *Top.* VIII 2, 158a25-28: «Whoever keeps on asking one thing for a long time is a bad inquirer. For if he does so though the person questioned keeps on answering the question, clearly he asks a large number of questions, or else *asks the same question a large number of times*: in the latter case he merely *babbles* (*ἄδολεσχεῖ*), in the former he fails to deduce». Protagoras would not be asking the same question many times, but would be proposing the same relativised claim many times, thus implicitly *asking* his opponents to concede it.

³ Could Protagoras say that their belief is *false*? Of course he could, but he would be thus unwittingly denying MD (cf. Dionysodorus' downfall in sect. 4. 1).

⁴ I agree with WATERLOW (1977, p. 27) that «171a6-c7 is not a proof of inconsistency». Some interpreters construe T9's argument as a logical proof by *Consequentia Mirabilis* (cf. sect. 6. 1): if MD is true, then it is false; therefore MD is false (cf. e.g. VAILATI 1904; KNEALE 1957, p. 63), overlooking T9's dialectical context.

4. 2. 5. One argument, two formulations

At the end of section 4.2.2 I suggested that the overall structure of the section 170a-171c has been generally misunderstood by commentators: it is time to illustrate more clearly my contention that T9's self-refutation argument must be interpreted as a clarification of T7's dilemma. I argued that T7's dilemma needed to be supplemented with missing qualifiers: although the addition was not too problematic in itself, it left us with an argument whose first premiss, (a*) («If men always judge what is true for them, then men sometimes judge what is false»), sounded unacceptable. T9 is designed to show Theodorus and us why (a*) is justified after all, by disclosing the «most clever» facet of Protagoras' MD:

| | 170c3-5 (a*) | 171a6-c4 |
|----------------------|---|--|
| MD | men always judge what is true <for them> (170c3) | all men judge what is the case <for them> (171a9) |
| Anti-MD consensus | (most) men believe that there are false judgements (since men believe in the existence of ignorance [170a6-b8] and (most) men believe that ignorance is false judgement [170b9-c1]) | (some) men (Protagoras' opponents) believe that MD is false (171a7-8) |
| ∴ | <i>Inference left unexplained</i> | <i>Inference explained at 171a6-b9</i> |
| MD is false | men sometimes judge what is false (170c5) | man is not the measure of anything at all which he has not learnt (171c1-4) |

The antecedent and the consequent of (a*) (MD and its contradictory, respectively) match, in content if not in exact wording, the initial premiss and the conclusion of T9's argument. What might lead us to suppose that the mysterious reasoning underlying (a*) cannot be the same as the one we find in T9 is that their additional premisses seem to differ: (a*) relies on the widespread *general* belief in the existence of false judgement, whereas T9's argument seems to appeal to the anti-Protagoreans' *specific* judgement that MD is false. However, this surface difference fades and almost disappears altogether as soon as one considers more carefully the nature of the anti-MD consensus in T9: although «the opinion of those who judge the opposite (by which they think that he [sc. Protagoras] says the false)» suggests an explicitly verbalised and specifically directed dissension against Protagoras, important clues show that the anti-MD consensus need not assume such a definite form. At 170e8-9 we are told that the *masses do not believe* MD, where the context makes it quite clear that this is intended as meaning that they *believe* MD *to be false*, and at 171b10-11 we find the bold conclusion that, since Protagoras is compelled to admit the falsehood of his MD, *everyone disputes it*. Since it is plausible that in the Athens of the beginning of the fourth century B.C. the vast majority of people never heard of Protagorean relativism, and, *a fortiori*, so much the less bothered disputing it actively, Socrates must be adopting a generous notion of belief here: you can be said to disbelieve, and even dispute, MD even if you have never heard of it, as long as you believe in the existence of false beliefs, which is inconsistent with MD.¹ Furthermore, this belief in the existence of false beliefs does not need to be ex-

¹ Pace NEWMAN 1982, p. 49.

plicit itself: you can be said to have it even if you have never expressed it,¹ or you do not have any full-blown concept of belief at all, provided that your behaviour *shows* this belief of yours. As we have seen above (p. 28), Socrates' own reasons for attributing to people the belief in the existence of ignorance and false judgement are based on the observation of their behaviour (in particular of their search for experts and leaders). Since the belief that MD is false is a corollary of the belief that false beliefs exist, T9's argument is nothing less than the expected elucidation of why (a*) is true and thus Socrates' initial dilemma is sound: it is meant to satisfy Theodorus' request for clarification at 170d3.²

A notable consequence follows from this analysis. T9's self-refutation argument is *dialectical*: it does not aim at proving that MD is false, but at showing why even Protagoras is bound to admit the falsehood of his MD when faced by someone who disagrees (or would disagree if asked), and so why no one can really uphold MD. Accordingly, T7's dilemma must also be interpreted as a dialectical challenge to Protagoras («What then, Protagoras, are we to make of your λόγος?»): whether he *says* (πότερον [...] φῶμεν) that MD is true, or that it is false, he will be finally compelled to *admit* that it is false, that «men do not always judge what is true <for them>». Socrates' dilemma is not a proof by cases of the necessary falsehood of MD: if this should appear an unwelcome result of my interpretation, it will be helpful to recall that Socrates' declared aim was to obtain *Protagoras' own agreement* starting from his own λόγος (169e8e-170a1), and not to demonstrate the falsehood of MD.

4. 2. 6. Protagoras' return

Faced with T10's unfortunate conclusion, Theodorus, who was supposed to defend his dead friend (168e7-169a1) but could not help conceding all the steps which led to Protagoras' dialectical rout, protests (171c8-9). Socrates' reply is worth examining:

T11 It's likely, then, that he [sc. Protagoras], being older, is wiser than us; and if he suddenly popped up here from below, as far as the neck, he would probably accuse (ἐλέγξας) me of talking a great deal of nonsense, and you of agreeing with it, and then he would duck down again rushing off (καταδύς ἂν οἴχοιτο ἀποτρέχων). But I think we have to take ourselves as we are, and always say what seems to us to be the case (τὰ δοκοῦντα εἰς ταῦτα λέγειν). (171c11-d5)

This vivid image has attracted some attention in the literature, concerning not only its pictorial details,³ but also its philosophical significance. What would Protagoras say should he come back from Hades? And why would he rush off immediately afterwards? According to Burnyeat, Protagoras might try to defend himself by insisting that he presented MD not as an absolute truth, but as something *true for himself alone*. By this solipsistic move Protagoras would be refusing to enter fully into a common world with his opponents for genuine discussion, and Plato's image would represent this attitude: «coming from and retreating to another world from ours, he [sc. Protagoras] does not really leave the underworld» (1976b, p. 193n23). However, this would look more like a way of *defending* MD by clarifying and *narrowing* it than a way of *accusing* Socrates

¹ Cf. KETCHUM 1992, p. 78.

² Theodorus seems to understand this new formulation of the argument without difficulty.

³ Cf. BURNYEAT 1976b, p. 192n23; FORD 1994.

of talking nonsense. In a similar vein, Lee conjectures that Protagoras would try to escape the «exquisite argument» by restoring the qualifiers that Socrates deliberately and ‘unjustly’ left out, but «at the high price of showing us that he himself is not asserting anything we can or should take “seriously”» (1973, p. 248),¹ and thus he would be «no better than a vegetable» sprouting up from the ground (p. 252). This sounds implausible: first, a plant could hardly be depicted as running away; second, like the imaginary opponents of Aristotle in *Metaphysics* Γ and unlike the returned Protagoras, plants do not speak at all.²

I suggest a more literal interpretation of T₁₁: by attacking and trying to refute (ἐλέγξας) Socrates, presumably protesting that his argument was unsound,³ Protagoras is at the same time refuting himself, because his MD is inconsistent with the possibility of someone being mistaken and someone else proving that he says something false.⁴ Protagoras is trying to do (it does not matter whether successfully) something which, according to his own doctrine, is impossible, thus betraying the fact that he himself, at the end of the day, does not really believe and ‘live’ his MD. Protagoras does not *choose* to rush off, but *must* rush off: his ducking down can be interpreted as the metaphorical counterpart of the further reversal which he involuntarily incurs by attacking Socrates.

As Socrates had previously remarked, MD is utterly incompatible with dialectic and refutation⁵ (just as it is incompatible with teaching, at least as ordinarily intended):⁶

T₁₂ I say nothing about my own case and my art of midwifery, and how much ridicule we incur; and I think the same goes for the whole business of dialectic (διαλέγεσθαι). For mustn’t it be a long and enormous nonsense to examine and try to refute (ἐλέγχειν) one another

¹ For similar views of what qualifying MD as true only for Protagoras would amount to cf. the literature listed on p. 33n4.

² For analysis of other interpretations of T₁₁ cf. CASTAGNOLI 2004a, p. 30n146.

Cf. also BEMELMANS 2002, p. 83: «Socrates alludes in this passage to the option he did not choose for Protagoras in the self-refutation argument [...]. The self-refutation would not have succeeded, if Socrates had let Protagoras hold on to qualification. With this choice, however, Protagoras would have deprived the others of their status of measure».

According to McCabe, the image is meant instead to express the doubt «whether Protagoras – who has Heraclitean leanings – could persist long enough, or with enough concreteness, to engage in conversation» and defend himself (2000, p. 47). I do not see any generic question of personal identity at stake here; there seems to exist a direct link between what the returned Protagoras says and *his* sudden ducking down (Protagoras does not dissolve while rushing off).

Ford argues that Protagoras’ only possible reply is simply to restate his *kephalaion*, the head statement of his book: «Protagoras’ head [...] runs off because it will still survive as a kind of *phēmē* (saying). Dismissed but not obliterated, the dead thinker’s saying is so well known, even notorious, that it is likely to pop up elsewhere among others interested in philosophy, and vex them with the same difficult words» (1994, pp. 204–205). This reading also ignores the point that Protagoras’ head is supposed to try to refute Socrates and Theodorus, and not simply to restate his position.

Other commentators do not try to conjecture what Protagoras’ rejoinder might have been, but suppose Protagoras rushes off because he has no good point to make (cf. e.g. CORNFORD 1935, p. 80; POLANSKY 1992, p. 132; EMILSSON 1994, p. 144n14; LEE 2005, p. 56, who speaks of «sterile stubbornness»). McDOWELL (1973, p. 171) believes that the image of Protagoras’ return could indicate that Plato is not content with his argument (possibly because of the lack of qualifiers). NARCY (1986, p. 81) too believes that the image indicates Plato’s dissatisfaction with his argument, but for different reasons (cf. p. 41n1).

³ The same kind of complaint had already been voiced, not without irony, at 166a–c, where Protagoras had lamented Socrates’ use of unfair verbal traps. Protagoras’ return had been foreshadowed at 169c8–d1.

⁴ Cf. WATERLOW 1977, pp. 28–29; GOTTLIEB 1992, p. 190. A similar interpretation is proposed but discarded by BURNYEAT (1976b, p. 191).

⁵ This important point is nicely stressed by LONG 2004.

⁶ Cf. *Th.* 161d8–e1.

er's appearances and judgements, when everyone's are correct if Protagoras' truth is true [...].¹ (161e4-162a2)

It should come as no surprise then that MD is dialectically untenable: any attempt to defend it inevitably results in an involuntary admission of its falsehood. This makes T₁₁, rather than T₉,² very similar to the *Euthydemus* passage of section 4.1: the οὐκ ἔστιν ἀντιλέγειν thesis was considered self-refuting in a dialectical context because its defenders were obliged, by the nature itself of that context, to try to throw their opponents' arguments down, thereby overthrowing their own position too, by implicitly conceding that false judgement and contradiction do exist. Both in the *Theaetetus* and in the *Euthydemus* it is absolutely immaterial whether Socrates' attacks are backed by sound arguments or poor reasoning:³ as Socrates mockingly reminds us in T₁₁, it is Protagoras himself who guarantees his adversaries the right of always saying whatever passes through their mind. In the light of these considerations, Burnyeat's motivation for discarding the qualifiers⁴ weakens: by such a move Socrates would not be displaying «perverse dishonesty», but showing how hopeless Protagoras' MD turns out to be as soon as submitted to dialectical scrutiny. Either Protagoras remains silent, or he angrily protests that Socrates is *misrepresenting* his doctrine and proposing an *incorrect* refutation (thus depriving Socrates of the status of measure as far as MD's exegesis and logical soundness are concerned), thus contradicting MD itself:⁵ in either case, Protagoras is the loser, because he cannot 'downthrow' adversaries and arguments which anyone else could have charged with «perverse dishonesty».⁶ Even an apparently shameless move would establish a powerful philosophical point: although I have provided a reconstruction of Socrates' argument which does not appeal to the absence of qualifiers, one might suppose that Plato's undeniable ambiguity, in particular in the elliptical T₇, is deliberate.⁷ Also an irrelevant charge can be lethal for one endorsing Protagoras' position.⁸

¹ Socrates fails to consider the possibility that Protagoras chooses a radically different form of dialectic, a 'therapeutic dialectic' in which the sophist argues to *change* his interlocutor's beliefs not from false to true, but from harmful to beneficial (cf. p. 27). According to NARCY (1986, pp. 80-81; 1995, pp. 93-100) the self-refutation argument is weak and 'eristic' exactly because it fails to consider this possibility. Notice, however, that this line of defence could be countered by Socrates' later argument for the non-relativity of expert predictions about the future and what is beneficial (177d-179b). ² Cf. p. 24n2.

³ Just as it is immaterial whether Protagoras' and Euthydemus' responses, respectively, really undermine Socrates' arguments or only purport to undermine them. ⁴ Cf. p. 29.

⁵ Cf. the dialectical manoeuvre described by Aristotle at *SE* 15, 174a20-23, in a chapter in which he lists various methods which help one to refute one's adversary more easily: «Moreover, there are anger and contentiousness, for when agitated everybody is less able to be on his guard; elementary rules for producing anger are to make it clear that one wishes to be *unfair* and to be completely *shameless*».

⁶ It is then quite ironic that Protagoras' *Truth* came to be known in antiquity also under the title *The downthrowers* (cf. p. 23n1). The use of the wrestling metaphor for dialectical argumentation is widespread also in the *Theaetetus* (cf. e.g. 162b, 166b1, 167e6, 169a-b).

⁷ Along these lines, one might argue that Socrates insisted that Theodorus' participated, in place of Theaetetus, to the examination of MD (168c-169c) because of his awareness that Theodorus, unlike the philosophically more gifted Theaetetus, would fail to notice the fallacious drop of the qualifiers, thus compelling Protagoras to return and defend himself *in propria persona*, with the disastrous consequences just explained. Notice that Theaetetus had already been made aware of the importance of dealing carefully with qualifiers in his discussion of previous objections to Protagoras (cf. p. 29n2), especially at 165b-c. For the related but different idea that the relativisers can be dropped in the self-refutation argument because Theodorus is Socrates' interlocutor, and «a sober mathematician would be the last person we would expect to take exception to the use of unrelativised terms» cf. LONG 2004, p. 36.

⁸ In the light of what we have seen in this section, I cannot agree with Barnes' comment that in Plato «the dialogue form is extrinsic in this sense: Plato's arguments can *all* be turned into monologues without any logical or philosophical loss» (BARNES 2003, p. 28, italics mine).

5. ARISTOTLE: SPEAKING TO ANTIPHASIS

The most conspicuous feature of Aristotle's attitude towards self-refutation is probably his apparently scarce interest: in nearly 1,500 Bekker pages of Greek prose often stuffed with dense argumentation, only a bunch of serious candidates to the role of self-refutation arguments leap to the eye. Nevertheless, since most of them target theses belonging to the broad family we are currently investigating, our inquiry will not lack an Aristotelian section.

5. 1. *Metaph.* Γ 4, 1008a27-30

In the war he wages in *Metaphysics* Γ against whoever denies PNC (hereafter, also «Antiphasis»),¹ starting from 1007b18 Aristotle focuses on the thesis that «the contradictories are all simultaneously true of the same thing», i.e. «anything affirmed may also be denied» and «anything denied may equally be affirmed» (1008a12-13):

$$(1) (\forall x)(\forall P)(Px \leftrightarrow \neg Px).^2$$

Having argued that this thesis commits its supporters to the view that «everything is one» (1007b18-1008a2) and to a breakdown of the principle of excluded middle (1008a2-7), Aristotle considers the issue whether, according to it, one could truly state the two members of the contradictory pair separately or only their conjunction, and shows that in either case unacceptable consequences follow. Aristotle depicts the unwelcome consequences of the first option as follows:

T13 Equally, even if it is possible to say the truth <in stating affirmations and denials> separately it follows [...] that (a) everyone will say the truth and everyone will say the false, and (b) he [sc. Antiphasis] himself admits ($\delta\mu\omicron\lambda\omicron\gamma\epsilon\tilde{\iota}$) that he is saying something false. (Γ 4, 1008a27-30)

Let us consider first consequence (a). Why if, for any predicate P and any thing x , one could *truly* affirm Px and truly deny it (i.e. assert $\neg Px$), would everyone say the truth and at the same time say the *false*? Would not everyone simply say the truth? Aristotle must be tacitly applying here the basic semantic principle according to which whenever an affirmation is true the corresponding denial must be false, and *vice versa*:

$$(2) (\forall x)(\forall P)(TPx \leftrightarrow F\neg Px).$$

Suppose that both Pn and $\neg Pn$ are true; Pn will be true, by hypothesis, and at the same time false, because of the truth of its contradictory, $\neg Pn$ (for (2) and double negation). On the basis of this simple reasoning, it is easy to see that if «the contradictories are all simultaneously true of the same thing» whatever one can say will be both true *and false*, and thus everyone will be both right *and wrong* about everything. This entire train of reasoning is crystal-clear: is it also sound? (2), which is the real, albeit implicit, pivot of Aristotle's argument in T13(a), looks unimpeachable: far from being a contentious prin-

¹ Cf. DANCY 1975. For in-depth discussion of Aristotle's views on contradiction, PNC and related principles cf. CAVINI 2007 and 2008.

² (1), together with the principle of excluded middle, entails $(\forall x)(\forall P)(Px \wedge \neg Px)$, i.e. the *contrary* (and not simply the contradictory) of PNC. For discussion of this 'curious turn' in the argument of Γ cf. WEDIN 2003.

ciple, it can be described as a corollary of the classical definition of truth and falsehood which Aristotle himself will provide in Γ 7:

T14 For to say that that which is is not or that which is not is, is false; and to say that that which is is and that which is not is not, is true. (1011b26-27)

If Pn is true, it must be so because it says that n (which is P) is P ; but then $\neg Pn$ must be false, since it says that n (which is P) is not P .¹

However unproblematic (2) might be *in itself*, I suggest that Aristotle's use of it could be questioned in this context. One who supposes that all contradictories are simultaneously true, like Antiphrasis here, is thereby likely to be asking us, if implicitly, to bracket the validity of a principle like (2). Otherwise, it would have been reasonable for him to claim that all affirmations and denials are at the same time true *and false*, and charity demands that we attribute him a position which is at least *prima facie* internally consistent. Given what I have explained about the intimate link between (2) and Aristotle's definition of true and false in T14, we must suppose that Antiphrasis is also asking us to revise radically our (for him) misguided conception of these notions. One might then argue that (2) cannot be tacitly presupposed in any argument against him, *insofar as such an argument aims at being dialectical* (as T13's final clause $\alpha\upsilon\tau\acute{o}\varsigma \alpha\upsilon\tau\acute{o}\nu \delta\mu\omicron\lambda\omicron\gamma\epsilon\acute{\iota} \psi\epsilon\upsilon\delta\epsilon\sigma\theta\alpha\iota$ suggests), and thus the opponent's defeat should be a consequence of his own thesis alone, or at least of views upon which he has agreed or would certainly agree. Aristotle might be accused of begging the question and allowing his *petitio principii* to sneak unnoticed into his argument in T13(a) in the form of (2).

That Aristotle could hardly have failed to be aware of such a possible rejoinder is testified by a passage occurring only a few lines below T13:

T15 if whenever the affirmation is true the denial is false, and when the latter is true the affirmation is false, there can be no such thing as simultaneously affirming and denying the same thing truly. However, they would probably assert that this is the issue originally posed. (Γ 4, 1008a34-b2)

Here Aristotle formulates the principle which I have labelled (2), and claims that PNC can be inferred from it.² However, he comments that the deniers of PNC would probably reject his argument and complain that it begs the question, because (2), in a sense, is nothing else than what Aristotle posed as the thesis he wanted to defend, PNC itself.

¹ As Walter Cavini suggested me, one might object that, if *taken literally*, T14's «definition» actually expresses only sufficient conditions for (saying the) true and (saying the) false, i.e. the rules of «semantic ascent»

$$\begin{array}{ll} Pn \rightarrow TPn & \neg Pn \rightarrow T\neg Pn \\ Pn \rightarrow F\neg Pn & \neg Pn \rightarrow FPn \end{array}$$

and not *necessary and sufficient* conditions (i.e. equivalences), and that the corresponding rules of «semantic descent»

$$\begin{array}{ll} TPn \rightarrow Pn & T\neg Pn \rightarrow \neg Pn \\ FPn \rightarrow \neg Pn & F\neg Pn \rightarrow Pn \end{array}$$

and thereby the equivalences can be extracted only from *Cat.* 12, 14b15-20. It should be noticed, however, that at *Metaph.* Γ 8, 1012b5-11 Aristotle himself seems to treat a variant of (2) as fully equivalent to his previous account of «what “false” and “true” signify». For a recent analysis of T14 with extensive bibliographical references cf. CRIVELLI 2004b, pp. 132-136.

² I do not see any reason for interpreting T15 as presenting an *indirect* argument, along Wedin's proposal (2000, pp. 160-161): if PNC were false, then (2) would be false, but this is absurd, and thus the denial of PNC is absurd too. I see even less reasons for describing the passage as arguing that «the argument may be self-refuting» (p. 159) or that «the argument may be self-defeating» (p. 162) (where «the argument» is, I suppose, Antiphrasis' denial of PNC).

To be precise, Aristotle does not admit that his opponents would be justified in complaining, but his hidden reliance on (2) in T₁₃ could seem liable to the same charge as his explicit assumption of it in T₁₅.¹ The fact that, to my knowledge, no commentator, apart from Alexander of Aphrodisias, noticed Aristotle's tacit, and possibly question-begging, manoeuvre in T₁₃² shows how difficult it is to identify certain unstated assumptions when these are so deeply rooted in our ordinary practice of inference, and actually constitutive of it. These are assumptions which usually we do not need to make explicit and we do not challenge, but in *Metaphysics* Γ it is some such assumptions and their consequences that face Antiphrasis' challenge, and so what is not question-begging in other contexts risks becoming so here.

Also T₁₃(b), i.e. Antiphrasis' self-refutation by the admission that his own thesis is false, crucially relies on the application of (2). First, the supporter of the thesis that whenever the affirmation is true the denial is also true will be forced to agree that the denial of his thesis is true as well.³ But then, as Alexander remarks,

T16 by his own admission that the denial of the affirmation which he himself posited [...] is true, he admits that he himself says what is false (αὐτὸν ὁμολογεῖ ψεύδεσθαι). (in *Metaph.* 296, 19-21)

The final step of the argument is not the one based on the self-application of Antiphrasis' thesis; it relies, once again, on the further application of (2) to the result of that self-application. In the absence of this extra step, Antiphrasis, who defends the contradictory⁴ of PNC, could be easily led to admit the truth of PNC *too*: he could thus be accused of having *omitted* some relevant truth, but not of having affirmed something he himself must admit to be *false*.⁵ Emphasising that Antiphrasis, the champion of contradiction, is committed to contradictory positions also about PNC is not necessarily a knock-out blow. Lear noticed that Aristotle's arguments in *Metaphysics* Γ are

constructed so as to reveal *to us* that Aristotle's opponent is in a contradictory position. [...] Aristotle is not trying to persuade him [*sc.* Antiphrasis]: the argument is for our sake, not for his. (Lear 1980, p. 113)

While agreeing on the essence of Lear's comment, I signal that it overlooks two important details. First, Aristotle believes that those who deny PNC as a result of philosophical confusion (caused, for example, by the flux of perceptibles or conflicting ap-

¹ Subsequently Aristotle shows no qualms about arguing, in a similar vein, for the principle of excluded middle starting from T₁₄'s definition of truth and falsehood (Γ 7, 1011b25-29) and against the thesis that everything is false starting from a variant of (2) (Γ 8, 1012b5-11). These different attitudes might depend on the different *degrees of similarity* between the disguised question-begging premiss and the conclusion, on the presence and number of additional premisses involved, and on the number of deductive steps needed to remove the disguise (cf. DANCY 1975, p. 20). For Aristotle's views on begging the question cf. *top.* VIII 13; *SE* 5, 6, 27; *APR* II 16, 64b33-38; see SCHREIBER 2003, pp. 98-106. The concern to avoid begging the question against Antiphrasis (or at least to avoid appearing to do so) by choosing a strategy, the 'elenctic proof', immune to this charge, was a priority in Aristotle's agenda from the very beginning of Γ 4.

² Alexander, commenting on T₁₅, remarks that Aristotle «has implicitly used this argument already, when he said that all were in error» (297, 11), clearly referring to T₁₃.

³ Aristotle does not take into account here the possibility that Antiphrasis might want to exempt his higher order generalisation from self-application («whenever the affirmation is true the denial is also true, *except for* this very affirmation, whose denial is not true»). For Aristotle's awareness of the possibility of such a move cf. the end of sect. 5. 2.

⁴ Or even the contrary of PNC, depending on the different occasions (cf. p. 42n2).

⁵ Moreover, there would be an answer available to him: «I omitted to state and defend PNC because, unlike its denial, it is something which is already so widely accepted, as you yourself suggest, that it does not need to be advocated by me or anyone else».

pearances) can and must be convinced, and doubtless a way of achieving this purpose consists in fully spelling out all the bizarre consequences of their denial (Γ 5, 1009a17-20). Perhaps the double consequence of T₁₃ is not the most impressive within Aristotle's overall enterprise of *Metaphysics* Γ 4-8, but it could still contribute to this pedagogical aim. Second, Lear undervalues the specific force of the admission that one's position is false. The fact that Antiphrasis should be ready to accept contradiction does not automatically imply that he will be happy to admit that his view is *false*. The extra step taken by Aristotle in T₁₃ thanks to his tacit application of (2) is not unimportant: one's admission that one is in error is much more embarrassing than one's admission that one's adversary is right too, and so much more so for a character like Antiphrasis, whose position has already been assimilated to Protagoras' at the beginning of Γ 5 and whose denial of PNC could therefore be read as a companion to the now familiar theses that everything is true and falsehood and error are impossible.

This is why Aristotle's strategy in T₁₃ is, at the end of the day, less airtight than it might appear: we are examining a peculiar case in which forcing one's adversary to concede the contradictory of his own thesis is not a sufficiently clear-cut victory, and Antiphrasis' stronger confession that he is in error can be obtained only by relying, surreptitiously, on an assumption which he might want to reject. But at what cost? Could he really be so bold to complain that Aristotle is guilty of *petitio principii* and refuse to grant (2)? So far I have assumed he could, and I confirm that this is theoretically correct. But, as I have already noticed, by rejecting (2) Antiphrasis would be rejecting, ultimately, that commonsense notion of truth and falsehood which is conveyed, for example, in T₁₄. If, on the one hand, Aristotle risks begging the question in his struggle with Antiphrasis, Antiphrasis, on the other, seems to be liable to the charge of *changing the subject* if he stubbornly refuses to grant at least some minimal features of our (and Aristotle's) semantic notions. Of course he is free to provide an alternative account for truth and falsehood; however, it cannot be *so* different from ours as to obscure the fact that he is trying to give an extremely revisionary account for those very things which we call «truth» and «falsehood».

Does Aristotle tacitly avail himself of (2) in T₁₃ because he is confident that his opponent could hardly dare reject it? Given Aristotle's caution in T₁₅, we cannot exclude that he simply hoped his move would pass unnoticed (what indeed happened).¹ However, this does not mean that Antiphrasis emerges in very good shape from T₁₃'s twofold attack: his only possible line of defence would make his thesis even more unpalatable, by requiring him to uncover and spell out certain consequences of his position on truth and falsehood which make it almost unintelligible.

5. 2. *Metaph.* Γ 8, 1012b13-22

Aristotle's most manifest adoption of the self-refutation charge against Antiphrasis occurs almost at the end of *Metaphysics* Γ:

T17 Indeed all such theses² are exposed to the stock objection (τὸ θρυλούμενον) that they eliminate themselves (αὐτοὺς ἐαυτοὺς ἀναιρεῖν). For anyone who says that everything is true al-

¹ Cf. however p. 44n1.

² Ross' (1924) translation of λόγους as «arguments» is incorrect: as Aristotle's subsequent explanation clarifies, it is the statements «Everything is true» and «Everything is false» that eliminate themselves (*contra* also CASSIN, NARCY 1989: «argumentations»; D'AGOSTINI 2002, p. 17: «ragionamento»).

so makes the thesis contrary to his own true, so that his own is not true (for the contrary thesis denies that his is true), and anyone who says that everything is false <makes> himself also <false>. And if they make exceptions – the former that only the contrary <of his own thesis> is not true, the latter that only his own is not false – they will end up none the less begging an infinite number of theses, true and false; for he who says that the true thesis is true says something true, and this will go on to infinity. (Γ 8, 1012b13–22)

Whereas T13's consequence (b), which I have reconstructed as a self-refutation charge, was not highlighted as such, Aristotle refers here to the strategy employed as *θρυλούμενον* («expressed over and over»), thus hinting at its previous history, and describes its thrust in wholly general terms: showing that certain *λόγοι* «eliminate» themselves (or, more tragically, «kill themselves», «are suicidal»). Aristotle's use of *θρυλούμενον* with reference to this charge captured Narcy's attention:

Platon, corrélativement, n'est pas nommé: Aristote n'estime pas que la mention de cet argument mérite plus que l'anonymat du lieu commun. Le terme dont il le désigne, *to thruloumenon* (1012 b 14), s'inscrit [...] dans le registre du bavardage [...]. Cette façon de mentionner l'argument platonicien n'est pas à proprement parler invalidante: d'un tel lieu commun il peut à l'occasion être commode se servir [...], et c'est pourquoi Aristote l'ajoute ici à son arsenal; mais elle est à coup sûr dépréciative. L'argument platonicien n'est rappelé que pour mémoire, en quelque sorte, moyen vulgaire de réfutation. (Narcy 1989, p. 81)

I agree that Aristotle shows no special enthusiasm for the two self-refutation arguments of T17.¹ Certainly their position and role within the overall architecture of *Metaphysics* Γ appear to be peripheral when compared with other arguments, in particular with the 'elenctic proof' of PNC in Γ 4. The self-refutation argument indisputably plays a more crucial part in Plato's *Theaetetus*, but there too it does not occur in a climactic position and I have argued that Plato himself, contrary to what is usually supposed, does nothing to highlight it as «especially clever». I suggest that, *pace* Narcy, Aristotle's way of introducing T17's self-refutation arguments does not necessarily reflect a veiled criticism of Plato's strategy.² Even if we understood *θρυλούμενον* in its most pejorative nuance («babbled over and over»), which is by no means necessary given Aristotle's own usage of the term,³ the fact that at the time in which Aristotle wrote Γ 8 the self-refutation objection seemed quite trite does not imply that he considered the achievement of those who first devised and used it trivial, or the force of the objection itself questionable. More importantly, calling an argument 'commonplace' and then going on to deploy it successfully and show that it thoroughly defeats one's opponent is not so much a way of devaluing the argument itself as a way of *increasing* the humiliation of that opponent. If your thesis is so hopeless that even a hackneyed argument is sufficient to subvert it, then your defeat is certainly less honourable than it would be if an ingenious, and possibly controversial, refutation were needed to beat you: «At least since the time of Democritus and Plato everyone knows that certain theses are suicidal,⁴ and you still wander around trying to sell them to us?» From a rhetorical point of view, Aristotle's use of *θρυλούμενον* is no less effective than the argument itself which he thus labels.

¹ Cf. also WEDIN (2005, p. 172), who speaks of Aristotle's «less sanguine attitude [sc. than Plato's] about the prospects for» self-refutation arguments.

² *Contra* also CASSIN 1993, p. 535.

³ Cf. *Ath.* 16.7.4; *HA* IX 13, 615b24; IX 37, 620b11; *GA* III 5, 756b6; *Metaph.* M 1, 1076a28; *Rhet.* II 21, 1395a10; III 7, 1408b2; III 14, 1415a2.

⁴ Cf. T20 in sect. 6. 1, p. 50.

So far I have discussed some clues to the Aristotelian attitude towards self-refutation emerging from T17; I shall focus now on what T17 teaches us about the logic of self-refutation. Once again, what is subject to self-elimination is not abstract propositions, but *statements* along with their *proponents* (λόγοις, ὁ λέγων). Aristotle is not saying that if everything were true then also «It is false that everything is true» would be true and, *therefore*, the proposition «Everything is true» is necessarily false: the final, crucial inference is missing.¹ Aristotle's reference to the asserter of this thesis would be irrelevant if his aim were to prove something about its propositional content. What Aristotle claims is that whoever *says* that everything is true *makes* the contrary of his thesis true *too*: since asserting something is not a sufficient condition for making it true (except for a few special cases which are not at stake presently), Aristotle must mean that whoever says that everything is true *is thereby committing himself* also to the truth of the contrary of his own thesis,² and thus can be forced to admit the falsehood of his own («so that his own is not true»). The nature of the argument as a purely dialectical silencer of one's opponent is confirmed by the case of «Everything is false»: he who says that everything is false «makes himself also false», i.e. unwittingly concedes, *by self-application*,³ that what he is saying must be false too (again, the further conclusion «Therefore it is false that everything is false» is missing). The proposal that nothing is true is thus self-defeating and suicidal: this is all that Aristotle seems to be interested in establishing through the self-elimination 'stock' objection. This is not to say that Aristotle did not believe that the propositions «Everything is true» and «Everything is false» are false, or that he thought he could not prove (at least in some weak sense of 'proving') their falsehood, but only that this is not what T17 purports to establish. This is a welcome discovery, since, as we have learnt, without first considering what the truth-bearers are and what their truth-value is,⁴ all one might prove is that «Everything is false» is either false or *liar-paradoxical*.

One final noteworthy aspect of T17 is Aristotle's mention and criticism of the attempt to elude the self-elimination charge by making some *exceptions*, i.e. by limiting the scope of the universal «everything». To my knowledge, this is the first discussion of this strategy to be found in the ancient texts: what was strikingly absent from the *Theaetetus* was precisely the suggestion that Protagoras might try to defend himself by claiming that man is the measure of all things, *apart from* man's being or not being a measure.⁵

¹ Pace D'AGOSTINI 2002, p. 17. This extra inference would be required to have what in logical jargon is often called *Consequentia Mirabilis*. I shall argue in sect. 6. 1 that *Consequentia Mirabilis* might have been unavailable to Aristotle.

² It is not clear whether the *contrary* (ἐναντίον) of his own thesis is supposed to be «Something is false», «Everything is false», or, more probably, given Aristotle's own wording, «It is not true that everything is true».

³ From this point of view, the present argument differs from Mackie's proof based on T-prefixability, *pace* DE PRAETERE (1993, p. 356).

⁴ For detailed discussion of Aristotle's complex position on the bearers of truth and falsehood (including, among the others, sentence-tokens, thought-tokens, and perhaps states of affairs and sentence-types) cf. NUCHELMANS 1973, pp. 23-44 and CRIVELLI 2004b, pp. 45-76.

⁵ *Contra* SEDLEY 2004, p. 48. Aristotle's criticism of this manoeuvre is based on the infinite repeatability of the true and false predicates, which would commit the deniers of truth and falsehood to conceding anyway infinitely many truths and falsehoods respectively. For the analysis of the 'exception move' in two later self-refutation contexts cf. BURNYEAT 1997, CASTAGNOLI 2000, pp. 303-306.

5. 3. *Metaph.* K 5, 1062a36-b11

Immediately before the passage we shall analyse in this section, the author of K¹ remarks that although there can be no proper proof ($\acute{\alpha}\pi\acute{o}\delta\epsilon\iota\chi\iota\varsigma$) of PNC and similar principles, an *ad hominem* ($\pi\rho\acute{o}\varsigma\ \tau\acute{o}\nu\delta\epsilon$) proof against those who deny them is possible (1062a30-31): for example, if one had questioned ($\epsilon\rho\omega\tau\acute{\omega}\nu$) Heraclitus in the right way, one might have forced him to admit ($\delta\mu\omicron\lambda\omicron\gamma\epsilon\tilde{\iota}\nu$) the truth of PNC, since he had adopted his opinion without really understanding what he was saying (1062a31-35).² We find then two arguments in a sequence. Let us begin by examining the first one:

T18 And in general if what is said by him [sc. Heraclitus] is true, not even this itself will be true, i.e. that the same thing can at one and the same time both be and not be. For as, when they are separated, the affirmation is no more true than the denial, in the same way – their combination and conjunction being like a single affirmation – the whole thing put forward as an affirmation will be no more true than its denial. (K 5, 1062a36-b7)

Unlike the other cases we have examined so far, no dialectical context is obviously presupposed here: what T18 does say is that if Heraclitus' thesis is true, then it is not true, and *not* that if one endorses the truth of Heraclitus' thesis, then one must admit it to be false. Surely we have found, finally, an unambiguous ancient example of a proof, by self-refutation, of the logical falsehood of a proposition? Such a diagnosis would be hasty. We should not overlook T18's broader context, and in particular Aristotle's observation, only a few lines above, that if one had *questioned* Heraclitus in the right way one might have compelled him to *accept* PNC. Must the conditional «if what is said by him is true, not even this itself will be true» be understood as the pivotal premiss of an implicit constructive dilemma argument along the following lines

| | | |
|-------|--|-----------------------------|
| (1) | If Heraclitus' thesis is true, it is not true; | $p \rightarrow \neg p$ |
| <2> | <If Heraclitus' thesis is not true, it is not true;> | $\neg p \rightarrow \neg p$ |
| <3> | <Either Heraclitus' thesis is true or it is not true;> | $p \vee \neg p$ |
| <hr/> | | |
| <4> | <Therefore, Heraclitus' thesis is not true.> | $\neg p$ |

or, once again, as an elliptical reminder of the dialectically self-defeating nature of Heraclitus' position? Undeniably, neither the unproblematic premisses <2> and <3> nor, more importantly, the alleged conclusion <4> appear in the text.³

T18 is followed by what sounds like a second self-refutation argument:

T19 Furthermore, if it is not possible to affirm anything truly, this itself will be false, the assertion that there is no true affirmation. But if some true affirmation exists, this would refute what is said by those who raise such objections and utterly destroy dialectic ($\tau\acute{o}\ \delta\iota\alpha\lambda\epsilon\gamma\epsilon\sigma\theta\alpha\iota$). (1062b7-11)

³ There is some doubt as to the authorship and chronology of *Metaphysics* K (cf. e.g. JAEGER 1934, pp. 208-219; AUBENQUE 1983). I shall analyse some *Metaphysics* K arguments without any commitment to their authorship and chronology.

² This is not in accordance with the typical current usage of informal fallacies theories, in which an *ad hominem* argument is an attempt to impugn one's opponent's views or arguments by somehow denigrating his person, character, and motivations («abusive *ad hominem*»).

³ The latter fact prevents us also from construing T18's argument as a *Consequentia Mirabilis* of the form $(p \rightarrow \neg p) \rightarrow \neg p$. I shall argue in sect. 6. 1 that *Consequentia Mirabilis* might have been unavailable to Aristotle.

The comments just made on T18 hold good, *mutatis mutandis*, here. It is hard to decide whether the argument is best interpreted as a dialectical silencer of *Antiphrasis* or as a proof of the necessary falsehood of the propositional content of *Antiphrasis*' assertion that no affirmation is true. This sort of indeterminacy often occurs when one tries to cast arguments formulated in a natural language into some precise logical form. However, the context and especially T19's final remark about the incompatibility of *Antiphrasis*' 'objection' with the practice of dialectic make me inclined to favour, again, the first interpretative line. This tentative approval might also be backed by familiar logical considerations: it would be incorrect to conclude that the affirmation that there is no true affirmation must be false, solely on the basis of its reflexivity, for supposing that in fact no other affirmation were true, «No affirmation is true» would be liar-paradoxical, in a way similar to Epimenides' «All Cretans say the false», and thus it would be hasty to call it unreservedly «false». The only conclusion one can safely draw about the truth-value of «No affirmation is true» *without inspecting what affirmations there are and what their truth-value is*, once again, that «No affirmation is true» is *either false or liar-paradoxical*. Although we are not sufficiently informed about the Aristotelian attitude towards the Liar,¹ charity invites us to read T19 in a way which does not commit its author to an incorrect, or at best superficial, approach to the thorny issues raised by this paradox and its cognates.²

6. INTRODUCING περιτροπή: SEXTUS EMPIRICUS

In the previous sections we have analysed several ancient arguments denouncing certain extremist views on truth and falsehood as hopelessly self-defeating, and detected numerous interesting analogies in their underlying logic. The passages we shall consider in this section, all from Sextus Empiricus' *corpus*, both testify to the continuous survival, and indeed flourishing, of the same argumentative pattern over the five centuries which separate Aristotle and Sextus and reflect a more precise awareness of its distinctiveness, starting from the frequent adoption of a semi-technical vocabulary to label it: the verb περιτρέπειν and its substantival form περιτροπή.

¹ Crivelli argues that Aristotle does refer to the Liar at *SE* 25, 180a34-b7 and «attempts to solve it by assuming that someone uttering "I am speaking falsely" (or whatever sentence-type the paradox turns on) is neither speaking truly nor speaking falsely absolutely» (2004a, p. 61; 2004b, p. 31). If this were correct (but cf. now CAVINI 2007, pp. 127-128), my argument for not interpreting T19 as a logical proof would find some further corroboration (provided, of course, that the author of *K* is Aristotle or someone who shares the same views on the Liar).

² There are at least two other arguments in defence of PNC often referred to in the literature as Aristotelian self-refutations: the short argument at *K* 6, 1063b30-35 and the 'elenctic proof' of PNC in *Γ* 4. I shall not deal with them here because, as I have argued at length elsewhere (CASTAGNOLI 2005), these are not genuine examples of self-refutation (at least on a reasonably precise and narrow conception of what counts as self-refutation). At *K* 6, 1063b30-35 «All statements are true» and «All statements are false» are rejected because absurdly incompatible with the definitions of true and false of T14, and not because of self-refutation (*pace* [Alex.] in *metaph.* 658, 20-26; CHERNISS 1935, p. 87n364; BURNYEAT 1976a, p. 44; REALE 1993, vol. III, p. 531). In *Γ* 4, Aristotle's 'elenctic proof' appears to have the structure of a *direct* refutation of not-PNC, where, importantly, *Antiphrasis*' thesis, not-PNC, does *not* figure among the premisses of its own refutation (cf. e.g. WEDIN 2000, p. 129), so not-PNC is not disproved by self-refutation (*pace* IRWIN 1988, CASSIN, NARCY 1989, BALTZLY 1999, POLITIS 2004, pp. 147-148). In a nutshell, the 'elenctic proof' of PNC is, unsurprisingly, an elenchus of *Antiphrasis*' not-PNC, based on certain *other* admissions. *Antiphrasis* cannot help granting if he wants to participate to *any* dialectical exchange (and not, narrowly, if he wants to state or defend his own denial of PNC).

6. 1. «Every appearance is true»:
dialectical reversal or Consequentia Mirabilis?

Let us begin with one of the best known and most discussed self-refutation arguments reported by Sextus:

T20 One could not say that every appearance (φαντασία) is true, because of περιτροπή, as Democritus and Plato taught contradicting Protagoras; for if every appearance is true, it will also be true, being based on an appearance, that not every appearance is true, and thus it will become false that every appearance is true. And even apart from περιτροπή of this kind... (M 7.389-390)

The noun περιτροπή («reversal», «about-turn») is used twice here to brand the kind of argument presented in the middle of the passage: what is this label supposed to mean and how does that argument work? In his first seminal article on ancient self-refutation, Burnyeat suggested that although «any refutation, of course, establishes the contrary of what it refutes», περιτρέπειν and περιτροπή tended «particularly to be used of the special case where the thesis to be refuted itself serves as a premise for its own refutation, where starting out with “*p*” we deduce “not-*p*” and so conclude that the original premise was false» (1976a, p. 48).¹ In Hellenistic and post-Hellenistic philosophical jargon περιτροπή would indicate not simply a reversal or refutation of a thesis into its contradictory, but a *self*-reversal or *self*-refutation having that thesis as its premiss.

Let us investigate how this broad outline fits the logic of T20's argument. If we take T20 at face value, we might infer that Democritus and Plato presented the following anti-Protagorean argument:

- | | | |
|-----|---|---------------------------------|
| 1 | (1) Every appearance is true | Assumption (Protagoras' thesis) |
| 2 | (2) It is an appearance that not every appearance is true | Assumption |
| 1,2 | (3) It is true that not every appearance is true | from (1) and (2) |
| ? | (4) It is false that every appearance is true | ? ² |

I shall not deal with the question of how Protagoras' relativistic outlook could turn, in most of the following philosophical and doxographical tradition (starting as soon as Aristotle), into an 'infallibilist' thesis to the effect that every appearance is (unqualifiedly) true. The *Theaetetus* passages we have examined in section 4. 2 could have contributed to the process, giving some readers the false impression that the Protagorean thesis under fire was an unqualified assertion that whatever appears is true, and Plato's complex argument might have been thereby read along the simplified lines recorded in T20. As for Democritus, although we do not possess any independent evidence for his employing (and possibly devising)³ any *such* argument against Protagoras, I see no serious reasons for doubting the reliability of Sextus' attribution (Plutarch informs us that Democritus wrote «many and persuasive things» against Protagoras).⁴ One could even

¹ I shall suggest that the final clause is questionable. It should be noticed that actually the fact that the self-refuting thesis «serves as a premiss for its own refutation» is not distinctive of self-refutation: any indirect proof, e.g. *reductio ad absurdum*, can be seen as a refutation of the contradictory of the conclusion having it as a premiss.

² I shall suggest that there are two alternative interpretations for this final step to be considered.

³ Sextus' order might indicate that Democritus used the argument *before* Plato, but cannot settle the issue. That Democritus used the argument before Plato is taken for granted, without argument, by Burnyeat (cf. p. 51n3). For Democritus' uncertain dates cf. LEE 2005, pp. 182-183.

⁴ *Adv. Colot.* 1109A4.

find a thin clue in favour of Sextus' accuracy in another Sextan passage, in which we are informed that Xenias of Corinth, «who was mentioned by Democritus too», asserted «that everything is false, and that every appearance and judgement are false» (M 7.53). If Democritus had something to say about such an obscure figure, we can perhaps suppose that it was to criticise the theses which Sextus attributes to Xenias.¹ One might thus conjecture that Democritus attacked Xenias' «Everything is false» by some self-refutation argument analogous to that which he employed against Protagoras' contrary position.²

What we can safely affirm is that, since the noun *περιτροπή* certainly did not appear in Plato, Sextus' own phrasing does not authorise us to believe that it was used by Democritus either: our source might be applying a later, Hellenistic label to an argument otherwise dubbed in both the authors he mentions. However, we should not even be too hasty in discarding the possibility that Democritus, unlike Plato, did use *περιτροπή*, or *περιτρέπειν*, and perhaps even prompted the philosophical career of these terms.³

But let us return now to the argument itself. It is clear that premiss (2) («It is an appearance that not every appearance is true») is necessary to infer, in conjunction with (1), the conclusion (3), that it is true that not every appearance is true. Burnyeat found this puzzling: what is the rationale of saying that (1) is *self-refuting*, if its reversal can be arrived at only with the aid of (2), which seems to be only contingently, albeit indisputably, true?⁴ One might protest that this perplexity arises from the ungrounded presupposition that in ancient *περιτροπή*, just like in modern absolute self-refutation, a *single* proposition must be involved: it is worth stressing that in fact in T20 not only we are not told that (1) is reversed by itself *alone*, but actually also the «by itself» is missing. Nonetheless, Burnyeat's solution deserves careful scrutiny:

We are to imagine Protagoras putting forward a subjectivist doctrine, according to which whatever appears to anyone to be so is so in fact, (1). He is opposed by someone saying that to him it appears, on the contrary, that not everything that appears to someone to be so is so in fact, (2).

¹ One could argue that actually Democritus might have had some sympathy for Xenias' second thesis, but only if «appearance» is interpreted in a narrow *perceptual* sense (cf. MOREL 1996, pp. 430-431). Cf. *Metaph.* Γ 5, 1009b11-12, where Aristotle claims that «Democritus asserts that there is no truth, or at least to us it is unclear» (but the context indicates that this report is meant to refer to sense-perception only) and *Cic. Luc.* 73, where Democritus is reported to have flatly denied that truth exists.

² Cf. T30 and T31 below. Democritus devised at least another famous self-refutation charge (fr. 125), with which we shall not be concerned here (cf. CASTAGNOLI 2005).

³ *Contra* BURNYEAT 1976a, p. 66: since «philosophical writing before Epicurus has plenty of occasion to speak of self-refutation, but the varied vocabulary used for the purpose makes no mention of reversal, [...] the idea of reversal can be traced to the first decade of the third century B.C. when the Hellenistic philosophies were taking shape». Burnyeat is perhaps too cautious by presupposing that it is *unlikely* that the *περιτροπή* vocabulary had originated well before Hellenistic times, with Democritus, on the sole basis that it is not to be found in Plato and Aristotle (after all, Democritus' name is conspicuously absent from Plato's writings, and as we have seen in sect. 5 Aristotle did not rely on self-refutation charges very frequently). One might suppose that Epicurus borrowed the *περιτροπή* jargon from Democritus and revived it.

Burnyeat seems to suggest that *φαντασία* is, just like *περιτροπή*, a later technical concept, extraneous to the original formulation of the argument (1976a, p. 47n5). This is not obvious: *φαντασία* is used twice in the *Theaetetus*, at 152c1 and 161e8, in the exposition of Protagoras' doctrine, both times with the relevant meaning of «what φαίνεται». Moreover, to insist that *φαντασία* is intended in its later technical sense in T20 could be dangerous for Burnyeat's own interpretation (see below): from a Stoic perspective, any mind content, even when not assented to but simply entertained, is a *φαντασία*, so the presence of a dialectical opponent risks becoming redundant, because the proponent of the thesis according to which every *φαντασία* is true must presumably entertain also the contradictory thought that not every *φαντασία* is true.

⁴ Cf. BURNYEAT 1976a, p. 49.

But Protagoras has only to be opposed like this and he will be forced to deny his own thesis and admit defeat, i.e. that not every appearance is true. His subjectivism is a non-starter, bound to lose him the debate before it has a chance to get going. It is necessarily a loser because in a dialectical context (2), contingent though it is, is in a sense guaranteed to hold; for there is no debate without disagreement and clash of views. [...] We began with the simple and, if you like, strictly self-refuting case of a thesis falsified by its own content. Then came falsification by the way a proposition is presented [...] Now it is the act of submitting a thesis for debate or maintaining it in the face of disagreement that causes its reversal and shows it up as false. One might call this dialectical self-refutation, and say that a thesis so falsified is dialectically self-refuting. (1976a, p. 59)

To have underscored the necessity of placing T20's *περιτροπή* in its dialectical background to do full justice to its logic was a fundamental contribution to our understanding of the argument; notice how similar the skeleton of the exchange reconstructed by Burnyeat is to that of Plato's own argument in the *Theaetetus* on the interpretation I defended in section 4.2. I have argued that *all* the ancient self-refutation arguments we have analysed so far are best understood as presupposing dialectical contexts, even when such contexts are not explicitly mentioned; Burnyeat suggests that the same approach might be required for T20 too. I can add here another clue to the existence of an implicit dialectical background in T20. According to Sextus, the *περιτροπή* argument shows why one could not *say* (οὐκ ἂν εἴποι τις), with Protagoras, that every appearance is true: while this could be the case in virtue of the fact that the self-refutation argument supposedly proves the falsehood of that thesis (and one should avoid saying something demonstrably false), it seems more natural to suppose that the reason is, simply, that if you do say that every appearance is false you will be easily defeated by your dialectical opponent in the way described.

This brings us to my first doubt about Burnyeat's conception of dialectical self-refutation, according to which dialectical self-refutation would «show up a thesis as false», or «falsify» it. If this were an alternative way of conveying the idea that the proponent of a dialectically self-refuting thesis is forced to deny it, or admit its falsehood, I would have no qualms about accepting such jargon. But since Burnyeat adopts the same vocabulary of falsification also for two other species of self-refutation (absolute and pragmatic), in which the falsification of the thesis involved is supposed to amount to an objective, non-dialectical proof of its falsehood, the notion of falsification is not innocuous here. The falsehood of the reversed thesis was indicated, in fact, as the final deductive step of *any* *περιτροπή* argument in Burnyeat's general account quoted above on p. 50 («and so conclude that the original premise was false»): it is not clear whether Burnyeat fails to distinguish falsification from dialectical reversal or supposes that the dialectical manoeuvre he describes is, or brings with itself, a falsification of the thesis involved as well. If we come back to T20's argument, however, we notice that I left its last step, (4), unexplained. From the Protagorean thesis, (1), and Protagoras' opponent's dissent, (2), Protagoras' unavoidable confession follows that (3) «It is true that not every appearance is true»; but then (4), «It is false that every appearance is true», is added as a further and final consequence. I suggest that the most natural way of reading this inference is to see (4) as a straightforward consequence of (3) ($T \neg p \rightarrow Fp$), an extra turn of the screw aimed at making Protagoras' defeat more glaring by having him admit explicitly the falsehood of his own thesis (and not only the truth of its contradictory). But, once again, Protagoras' own confession that (1) is false is not the same as a proof that (1) is false, nor does it imply it.

One might be tempted to propose a different interpretation of T20's argument as a non-dialectical proof of the logical falsehood of proposition (1), along the following lines:

| | | | |
|--------|-------|---|---------------------------------|
| 1 | (1) | Every appearance is true | Ass. (Protagoras' thesis) |
| 2 | (2) | It is an appearance that not every appearance is true | Assumption |
| 1,2 | (3) | It is true that not every appearance is true | from (1) and (2) |
| 1,2 | <3.1> | Not every appearance is true | from (3), by $Tp \rightarrow p$ |
| 2 | <3.2> | If every appearance is true, not every appearance is true | from (1) and <3.1> |
| 3.3 | <3.3> | If something implies its own contradictory, it is false | Law of classical calculus |
| 2, 3.3 | (4) | It is false that every appearance is true | from <3.2> and <3.3> |

I will not stress further the problem that, on such a reconstruction, (2) would be the assumption of a contingent extra-logical truth, which would remain 'undischarged' at the end.¹ It is the status of another assumption that appears to me no less problematic here: the conclusion, (4), would be a consequence ($\chi\alpha\lambda\ \omicron\upsilon\tau\omega$) not of (3), the apodosis of the preceding conditional sentence, but of the whole conditional, <3.2>, together with the implicit <3.3>. Now, <3.3> is a tautology of the classical propositional calculus

$$(CM) \quad (p \rightarrow \neg p) \rightarrow \neg p$$

often referred to by the specialists as *Consequentia Mirabilis*:² any proposition implying its own contradictory is false³ (precisely, necessarily false). CM has been lurking more than once in the footnotes in the previous sections,⁴ but I postponed the discussion of it to this stage of our inquiry for expository convenience.

I suggest there are various reasons for doubting the correctness of any reconstruction of T20's argument relying on CM. To begin with, and most obviously, no trace of CM appears in our text. To this objection, one might reply that CM does not need to be stated explicitly, because it is not a *premiss*, but an *inferential schema* ($p \rightarrow \neg p \vdash \neg p$), the

¹ See the left column in the derivation, where the assumptions on which each step depends are listed. Given the assumption (2), the argument does not suffer the same weakness as we encountered in previous 'logical' reconstructions of ancient self-refutation arguments: if it appears to someone that ($\neg p$) not every appearance is true, then it is in fact impossible that every appearance is true. For even if all the other appearances should be true, at least the appearance $\neg p$ will not be true, but Liar-paradoxical.

² Often the name *Consequentia Mirabilis* (attested for the first time in the 17th century among Polish Jesuit scholars: cf. ŁUKASIEWCZ 1970, p. 168n19) is primarily attributed to the formula

$$(\neg p \rightarrow p) \rightarrow p,$$

dubbed also *Lex Clavii*, and only consequently to CM above, which follows from it by a simple substitution of the variables and the law of double negation (for the *fortuna* of this label cf. BELLISSIMA, PAGLI 1996, pp. 205-212). I shall use CM both for $(p \rightarrow \neg p) \rightarrow \neg p$ and for $(\neg p \rightarrow p) \rightarrow p$ (notice, however, that the two formulas are not equivalent in all systems: in intuitionistic logic, for example, only $(p \rightarrow \neg p) \rightarrow \neg p$ is valid, whereas $(\neg p \rightarrow p) \rightarrow p$ is not). For a comprehensive history of *Consequentia Mirabilis* cf. NUCHELMANS 1991, pp. 124-137; BELLISSIMA, PAGLI 1996.

³ In this paraphrase of CM «is false» is used only as a convenient method of generalising about negation. However, if one accepts the basic semantic principle $\neg p \rightarrow Fp$, CM can be easily reformulated metalinguistically with a falsehood predicate: $(p \rightarrow \neg p) \rightarrow Fp$. To be precise, this reformulation seems to be the one required in the argument above to infer the conclusion (4):

| | | | |
|--------|-------|---|---------------------------------|
| 1 | (1) | $(\forall p)Ap \rightarrow Tp$ | Ass. (Protagoras' thesis) |
| 2 | (2) | $A(\neg(\forall p)Ap \rightarrow Tp)$ | Assumption |
| 1,2 | (3) | $T(\neg(\forall p)Ap \rightarrow Tp)$ | from (1) and (2) |
| 1,2 | <3.1> | $\neg(\forall p)Ap \rightarrow Tp$ | from (3), by $Tp \rightarrow p$ |
| 2 | <3.2> | $((\forall p)Ap \rightarrow Tp) \rightarrow (\neg(\forall p)Ap \rightarrow Tp)$ | from (1) and <3.1> |
| 3.3 | <3.3> | $(p \rightarrow \neg p) \rightarrow Fp$ | metalinguistic CM |
| 2, 3.3 | (4) | $F(\neg(\forall p)Ap \rightarrow Tp)$ | from <3.2> and <3.3> |

⁴ Cf. nn. 1 on p. 15, 4 on p. 37, 1 on p. 47, 3 on p. 48.

obvious validity of which would be recognised by any attentive reader as what justifies the inference from a conditional formed by contradictories, <3.2>, to its consequent, (4). However, not only is CM never *presented* as a 'logical law' by our sources for Hellenistic logic¹ or, more broadly, for ancient logic, but no argument is attested in which it is unequivocally *employed* (our T20 has been seen as one of the few and most promising texts for an attempt at tracing back the use, if not the theoretical formulation, of *Consequentia Mirabilis* to antiquity).² Although this might be partly imputed to an objective difficulty in univocally translating into logical form arguments expressed in a natural language, to say confidently that an ancient reader would have easily supplied the unstated CM may mean providing him anachronistically with a tool he would have been unable to handle. Moreover, the suggestion that CM is the underlying inferential schema which allows us to infer (4) from <3.2> is liable to another objection: deductions from a single premiss did not meet particular enthusiasm – to put it mildly – in antiquity, at least as far as we can judge from the two most influential ancient logical constructions, Aristotle's syllogistic and Chrysippus' dialectic.³

To avoid these difficulties, we could attempt a slightly different reconstruction, not involving CM, but a distinct, albeit strictly related, inferential schema. On this interpretation, after <3.1> the argument would proceed as follows:

- | | | | |
|-----|-------|---|---|
| 2 | <3.2> | If every appearance is true, not every appearance is true | from (1) and <3.1> |
| 3.3 | <3.3> | If not every appearance is true, not every appearance is true | Duplicated conditional |
| 3.4 | <3.4> | Either every appearance is true or not every appearance is true | Excluded middle |
| 2 | <3.5> | Not every appearance is true | from <3.2>, <3.3>, <3.4>, by constructive dilemma |
| 2 | (4) | It is false that every appearance is true | from <3.5>, by $\neg p \rightarrow Fp$ |

Although such a reconstruction would require us to supply even more implicit reasoning than before, which is sufficient to make it rather questionable as an exegesis of the text, the inferential pattern to the conclusion would be in accordance with a schema which, apparently, was recognised as sound and used in antiquity and which occurs elsewhere in Sextus' *corpus* (cf. p. 57n6):

¹ Bobzien puts the *Mirabilis* in a list of sequents that have such a form that no compounds of propositions of that form «would be syllogisms in the Stoic system, although all of them are correct sequents in $\neg c$ [sc. classical propositional calculus]». She claims also not to have found any documentation in the sources «that the Stoics accepted either all corresponding conditionals of a form as true, or a metalogical principle that in some way corresponds to the sequent» (BOBZIEN 1996, pp. 183-184). These sequents (including the *Mirabilis*) cannot be analysed into indemonstrables either.

² Commenting on T20, Bellissima and Pagli write: «Non c'è dubbio che lo schema di ragionamento, effettivamente parallelo a quello del *Teeteto* platonico [...], ripercorra le linee della *Consequentia Mirabilis* nella forma $(A \rightarrow \neg A) \rightarrow \neg A$ » (1996, p. 178). It should be clear from my reconstruction of the *Theaetetus* argument that I do not agree that it relies on CM either (cf. p. 37n4).

In sect. 5. 2 I have reconstructed Aristotle's treatment of the 'self-eliminating' statements «Everything is true» and «Everything is false» at *Metaph.* Γ 8, 1012b13-22 in a straightforward dialectical way which does not presuppose any unstated application of CM.

³ Aristotle's phrasing of the definition of συλλογισμός (cf. e.g. *APr.* 1 1, 24b18-21: « A συλλογισμός is a λόγος in which, certain *things* being stated, something other than what is stated follows of necessity from *their* being so») seems to exclude single-premiss deductions (cf. ALEX. APHR. in *APr.* 17,10-18,8); the Stoic Antipater is reported to have endorsed a very *unorthodox* (and much criticised) view by allowing single-premiss arguments (μονολήμματοι λόγοι).

| | | |
|-------|-----------------------------|-------|
| (a) | $p \rightarrow \neg p$ | <3.2> |
| (b) | $\neg p \rightarrow \neg p$ | <3.3> |
| (c) | $p \vee \neg p$ | <3.4> |
| <hr/> | | |
| (d) | $\neg p$ | <3.5> |

This form of constructive dilemma is strictly related to CM: it might be seen as an extended version of it, in which the two tautological extra premisses (b) and (c) are spelled out.

There is a final obstacle which such an exegetical approach to T20's argument must overcome: could the key premiss «If every appearance is true, not every appearance is true», if taken at face value,¹ ever be true? More generally, is it possible for *any* conditional of the form $p \rightarrow \neg p$ to be true? If one accepts the truth-conditions for 'material implication' typical of (modern) classical calculi, the answer is, of course, yes: $p \rightarrow \neg p$ is true whenever p is false. Those truth-conditions were accepted in antiquity by the dialectician Philo² and, most probably, by *some* Stoics.³ A different analysis, proposed by Diodorus Cronus,⁴ required the present impossibility of p (i.e., in Diodorean terms, the falsehood of p now and at any time from now on) for $p \rightarrow \neg p$ to be true. On both conceptions nothing would prevent <3.2> from being accepted as true. However, there is strong circumstantial evidence to make the case that two other particularly influential ancient conceptions of sound implication and conditional would have barred that possibility. In modern discussions of non-classical logics one radically non-classical thesis is sometimes singled out as distinctive:

$$(TA) \quad \neg(p \rightarrow \neg p) \quad \neg(\neg p \rightarrow p)$$

McCall, who propounded one of the first systems of «connexive logic», characterised by a brand of «connexive implication» such that «no proposition connexively implies or is implied by its own negation», baptised this property «Aristotle's thesis» (hereafter, also TA).⁵ This choice was not random, since according to McCall the following passage testifies Aristotle's endorsement of TA:

T21 But it is impossible that the same thing should be necessitated by the being and by the not-being of the same thing. I mean, for example, <that it is impossible> that B should necessarily be large if A is white and that B should necessarily be large if A is not white. [...] If then B is not large, A cannot be white [from «If A is white, necessarily B is large», by contraposition]. But if, if A is not white, it is necessary that B should be large, it necessarily results [by transitivity] that if B is not large, B itself is large. But *this* is impossible. (APr. II 4, 57b3-14)⁶

TA seems to be accepted at the very end of the passage and used by Aristotle to prove the impossibility of two implications of the form $p \rightarrow q$ and $\neg p \rightarrow q$ (baptised by McCall «Boethius' thesis», hereafter also TB):

¹ I.e. not as shorthand for «If every appearance is true and it appears that not every appearance is true, then not every appearance is true», which is a conditional of the different form $(p \wedge q) \rightarrow \neg p$.

² Cf. e.g. S.E. PH 2.110. A Philonian conditional, like a material implication, is false only when its antecedent is true and its consequent is false.

³ The adoption of Philo's truth-conditions by some Stoics emerges sometimes in our sources (cf. e.g. S.E. PH 2.104); the problem is to establish whether by pre- or post-Chrysippean Stoics, and how reliable these testimonies are.

⁴ Cf. e.g. S.E. PH 2.110-111.

⁵ McCall 1966, p. 415.

⁶ On this passage cf. Patzig 1959; Mignucci 1969, pp. 610-614.

| | |
|---|-----------------------------|
| Suppose that both (a) $p \rightarrow q$ and (b) $\neg p \rightarrow q$ | |
| If (a) $p \rightarrow q$, then (c) $\neg q \rightarrow \neg p$ | by contraposition (57b9-11) |
| If (c) $\neg q \rightarrow \neg p$ and (b) $\neg p \rightarrow q$, then (d) $\neg q \rightarrow q$ | by transitivity (57b6-9) |
| But (d) $\neg q \rightarrow q$ is impossible | TA |
| Therefore, $\neg(\text{both } p \rightarrow q \text{ and } \neg p \rightarrow q)$ | TB |

As William Kneale remarked, «if Aristotle was right in asserting this, there could never be any valid argument in the pattern of the *consequentia mirabilis*» (Kneale's bold conclusion was that Aristotle *was* certainly wrong, and «Aristotle's thesis» was indeed Aristotle's *error*).¹ No constructive dilemma in the form presented on p. 55 could be sound either, since its first premiss would be bound to be false.² More generally, *any* attempt at proving the necessary falsehood of a proposition (e.g. Protagoras' thesis) starting from the fact that it implies its own contradictory would be a non-starter.

One could object that T21 alone offers insufficient grounds for establishing Aristotle's real commitment to TA and TB;³ moreover, even conceding that T20's anti-Protagorean argument, in its non-dialectical reconstructions, perhaps would not appeal to Aristotelian devotees, it still could be convincing for readers with different logical tastes. In response to this, I suggest that the 'non-classical' attitude towards implication which apparently emerges from T21 was not isolated in the ancient logical landscape. There are compelling reasons to believe that the (most likely Chrysippean) conception of conditional (συννημμένον) dubbed συνάρτησις («connectedness») by some of our sources, which seems to have been Stoic orthodoxy for some time,⁴ also involved the truth of TA. Here is the definition of συνάρτησις reported by Sextus:

T22 Those who introduce συνάρτησις say that a conditional is sound when the contradictory of its consequent is in conflict (μάχεται) with its antecedent; according to them, the conditionals mentioned above [sc. «If it is day, I converse»⁵ and «If there are not indivisible elements of the things, there are indivisible elements of the things»]⁶ will be unsound, while «If it is day, it is day» will be true. (PH 2.111)

Admittedly, the attested falsehood of the conditional «If there are not indivisible elements of the things, there are indivisible elements of the things» on the συνάρτησις conception is insufficient to establish that *every* conditional of the same form $\neg p \rightarrow p$ must be false on that conception. However, another Sextan passage suggests the general validity of TA for συνάρτησις:

¹ «On this occasion Aristotle wrote more than was needed and fell into error» (KNEALE 1957, p. 66). ŁUKASIEWICZ (1957, pp. 48-51) and PATZIG (1959, p. 191) share the same opinion. For a more sympathetic approach, based on the request that the antecedent is explanatory of the truth of the consequent, cf. SMITH 1989, pp. 190-191.

² $\neg(\neg q \rightarrow q)$ entails $\neg(q \rightarrow \neg q)$ provided one accepts contraposition and double negation.

³ One could object that the second premiss of Aristotle's famous φιλοσοφητέον argument in the *Protrepticus*

| | | |
|-----|------------------------|---|
| (a) | $p \rightarrow p$ | If one must philosophise, one must philosophise |
| (b) | $\neg p \rightarrow p$ | If one must not philosophise, one must philosophise |
| (c) | p | Therefore, in any case one must philosophise |

is obviously incompatible with TA and that this suggests that we should not attribute TA to Aristotle. I have argued elsewhere at length (CASTAGNOLI 2005) that the argument in this form attested by our late sources should not be attributed to Aristotle himself, and that the original argument in the *Protrepticus* had a dialectical structure not reducible to CM and compatible with TA.

⁴ Most notably, the συνάρτησις truth-conditions for conditionals are the only ones presented by Diogenes Laertius in his testimony on Stoic logic (7.73).

⁵ True for Philo (PH 2.110).

⁶ True for Diodorus (PH 2.111).

T23 But it is impossible, according to what they¹ say, that a sound conditional is composed of conflicting propositions. (*PH* 2.189)

Since contradictory propositions are always in mutual conflict (*μάχη*) too, the thesis expressed in T23, which I shall call «Chrysippus' thesis» (hereafter, *TC*), sounds like strong evidence that *TA* conveys a genuine property of *συνάρτησις*. The reliability of this testimony, the importance of which was first underscored by Nasti,² has been questioned,³ but seems to have now found sufficient corroboration in other sources which can be thought to provide trustworthy, albeit indirect, information about Stoic logic. Although I cannot enter the details of this debate, which I have discussed at length elsewhere,⁴ let us consider at least a couple of passages:

T24 Where one can use the disjunctive connective (*ὁ διαζευκτικός*), one cannot use the conditional one (*ὁ συναπτικός*); and where one can use the conditional one, one cannot use the disjunctive one. And it is clear from what has been said that what the conditional <connectives> [...] announce is in conflict with what is announced by the disjunctive <ones> [...] between the disjuncts there is no consequence. (*Apoll. Dysc. de conj.* 218, 11-15)

T25 Moreover, the antecedents and the consequents cannot leave one another, nor can conflicting propositions (*repugnantia*) be connected (*cohaerere*) one to another; the former are necessarily connected one to another, the latter disconnected. (*Boeth. in Cic. top.* 349, 40-42)

Both Apollonius Dyscolus and Boethius report that a true conditional cannot be formed of conflicting clauses (*TC*), or, *a fortiori*, contradictory clauses (*TA*).⁵

There seems to be credible, albeit circumstantial, evidence, therefore, for attributing *TA* to at least some Stoics (Chrysippus and those endorsing his notion of *συνάρτησις*). If this attribution is correct,⁶ any reconstruction of T20's argument along the lines of

¹ «They» are «the dogmatists» (*PH* 2.111), but the context indicates that the reference is to philosophers who adopt the *συνάρτησις* truth-conditions.

² Cf. NASTI 1981.

³ Cf. STOPPER 1983.

⁴ Cf. CASTAGNOLI 2004b, 2004c, 2008. Cf. also NASTI 2002.

⁵ For further details about these passages and their significance cf. NASTI 2002; CASTAGNOLI 2004c, 2008; BARNES 2005. One might question the reliability of Apollonius Dyscolus and Boethius as sources for *Stoic logic*. Without tackling this objection directly here, I shall just reply that it need not be damaging for my present purposes: if *TA* appears in their reports but did not necessarily derive from Stoic logic, *TA* could have been endorsed even more widely. If *TA* holds good for *συνάρτησις*, given the ascending order of strength of the conceptions of conditional presented by Sextus at *PH* 2.110-112, the fourth one, *ἐμφάνσις* («manifestation»), should have been governed by *TA* as well.

⁶ Before we can accept *TA* as a genuine Stoic thesis, however, it is necessary to show how this can be reconciled with some attested dilemmatic arguments apparently incompatible with *TA* but usually attributed to the Stoics (cf. *PH* 2.131, 186; *M* 8.281, 466):

| | | |
|-------|------------------------|---|
| (a) | $p \rightarrow p$ | If cause (sign, proof) exists, cause (sign, proof) exists |
| (b) | $\neg p \rightarrow p$ | If cause (sign, proof) does not exist, cause (sign, proof) exists |
| (c) | $p \vee \neg p$ | Either cause (sign, proof) exists or cause (sign, proof) does not exist |
| <hr/> | | |
| (d) | p | Therefore, cause (sign, proof) exists |

I have argued at length elsewhere (CASTAGNOLI 2005 and 2008) that various textual clues allow us to reinterpret those dilemmas as *elliptical* and *enthymematic* formulations of dialectical arguments perfectly compatible with *TA* (and *TC*):

| | | |
|-----|--|---|
| (a) | $q \rightarrow p$ | If <you answer that> cause (sign, proof) exists, then <you yourself admit that> cause (sign, proof) exists |
| (b) | $r \rightarrow s \wedge s \rightarrow p$ | If <you answer that> cause (sign, proof) does not exist, then <you must present a cause (sign, proof) of what you assert, in order to be credible, but thus you yourself admit that> cause (sign, proof) exists |

CM or cognate schemas becomes unpalatable for a large audience (probably Aristotle, and part of the Peripatetic tradition, and mainstream Stoicism), and for serious logical reasons. If we add to this the indisputable datum that such a reconstruction fits Sextus' actual wording rather poorly anyway, I hope I have constructed a compelling case for sticking to a *purely* dialectical interpretation of T20's περιτροπή which does not require that Protagoras' thesis is proved to be 'logically false' because it entails its own contradictory. But, then, Burnyeat's presupposition that dialectically self-refuting propositions are (also) somehow falsified becomes dubious, and in the light of what I have explained about CM and TA his suggestion that, generally, in a περιτροπή argument «starting out with “p” we deduce “not-p” and so conclude that the original premise was false»,¹ which sounds like a description of CM, turns out to be suspect too.²

The absence of any distinction between the dialectical nature of περιτροπή and a 'logical law' like CM is a first aspect of Burnyeat's analysis which might require some revision. Another point of disagreement, which will emerge more clearly in the next section, concerns the existence, allowed by Burnyeat, of *some* περιτροπή arguments in which the dialectical aspect plays a limited role, or no role at all: as we have seen, dialectical περιτροπή is, according to Burnyeat, only one *species* of ancient self-refutation,³ along with others, like absolute and pragmatic self-refutation.⁴ I believe, instead, that all περιτροπή has a character not reducible to purely formal argument, without thereby denying that distinct species of ancient self-refutation can be usefully classified.⁵

Before passing to the inspection of some περιτροπή arguments allegedly amounting to absolute self-refutation,⁶ let us consider briefly three other Sextan passages which replicate and confirm some fundamental features which I have detected in T20:

T26 They will say (λέξουσιν) either that all things which appear are true, or that some are. If all, their λόγος incurs reversal (περιτρέπεται); for it appears to some people that nothing is true. (PH 2.88)

T27 Either all things which appear and all those unclear are true, or some things which appear and some unclear. Now if all are, their λόγος will again be reversed (περιτραπήσεται), it being granted as true (ἀληθοῦς διδομένου) also that nothing is true. (PH 2.91)

| | | |
|-----|------------|---|
| (c) | $q \vee r$ | Either <you answer that> cause (sign, proof) exists or <you answer that> cause (sign, proof) does not exist |
| (d) | p | In any case <you must admit that> cause (sign, proof) exists |

¹ BURNYEAT 1976a, p. 48 (italics mine).

² *Contra* also BARNES 1997, p. 31: «περιτροπή [...] turns on the exotic truth that anything which implies its own negation is itself false».

One could object that I have shown, at most, that 'logical' περιτροπή, intended as something akin to CM, was only available to some philosophical schools which did not endorse TA; however, Burnyeat himself maintains, I think correctly, that «evidently *peritrope* is a commonplace of later Greek controversy, *available to disputants of any persuasion* to confute the other side of the debate» (1976a, p. 57; italics mine).

³ In Burnyeat's own terms, it is an «extension of the notion of self-refutation» (1976a, p. 59), and not a correction or replacement of it.

⁴ In his analysis Burnyeat makes it clear that pragmatic περιτροπή often occurs in dialectical contexts, but the fact that «dialectical self-refutation» is chosen as the name of a species of περιτροπή risks obscuring this very important point.

⁵ Burnyeat juxtaposes, without stressing their differences, the general, non-dialectical account of περιτροπή which I have criticised above and the following dialectical account: «For precisely what self-refutation consists in is a reversal whereby advancing a proposal commits one to its contradictory opposite» (1976a, p. 49).

⁶ For reasons which should now be clear, I do not agree with McPherran's analysis of T20's argument as an instance of *absolute* self-refutation (1987, p. 293n8).

T28 And if the ‘something’ is true, everything will be true; from which it is in turn concluded that nothing is true, since this itself, I mean that nothing is true, being a thing, is true. (*PH* 2.86)

In T26 a dialectical exchange is unquestionably presupposed: *περιτροπή* is the result of one *saying* that all *φαινόμενα* are true, and thus what gets reversed, presumably into its own contradictory, is what one has said (*ὁ λόγος*).

In T27 the reason why the *λόγος* that all appearances are true incurs reversal seems to be, simply, that one who proposes it is thereby *conceding as true* also something which clearly amounts to its contradictory (that nothing is true). Nothing is suggested about the possible implications for the truth-value of that *λόγος*.

In T28 «Nothing is true» cannot be meant to be the true conclusion of a complete proof by self-refutation of the assumption that «Everything is true», but must be an unwelcome consequence of that undischarged assumption: if everything were true, then nothing would be true. Once again, no further step involving *CM* (e.g. «therefore, not everything is true») is presented or suggested. This time no dialectical framework is present either, but if one explores the broader context of T28 that framework can be easily supplied:

T29 Moreover, the ‘something’, which is, they [*sc.* the Stoics] say, the highest genus of all, is either true or false or neither false nor true or both false and true. If then they will say that it is false, they will be *admitting* that everything is false. [...] **T28** And if the ‘something’ is true, everything will be true...

Although at the beginning the four logical possibilities are listed in an abstract way, subsequently it is not the consequence of the first possibility that is analysed, but the consequence of the Stoics *accepting* it and *saying* that their *summum genus*, the ‘something’, is false. Symmetry requires us to understand the same also in the case of the second possibility in T28: «And if <the Stoics will say that> the ‘something’ is true, <they will be thereby admitting that> everything is true; from which it is in turn concluded that <they must concede that> nothing is true». We had already accepted this sort of ellipticity as a plausible explanation of certain otherwise problematic features of the *Dissoi logoi* argument (section 2) and of Socrates’ first dilemma against *MD* in the *Theaetetus* (section 4. 2. 2); more crucially, we shall appreciate soon that such conciseness is a feature of Sextus’ own *usus scribendi*.

6. 2. «Nothing is true»: two different approaches

I have already noted that according to Burnyeat some instances of *περιτροπή* («single-premise reversals») fit the description of absolute self-refutation: they exemplify the «strictly self-refuting case of a thesis falsified by its own content» (1976a, p. 59) through the pattern of a *Consequentia Mirabilis* of «the form “ $(p \rightarrow \text{not-}p) \rightarrow \text{not-}p$ ”» (p. 49). I shall examine now two of these arguments in Sextus, to verify whether the modern conception of absolute self-refutation, both in its bare bones, as just described, and in the fuller details of Mackie’s influential formalisation, really captures their logic.¹ Let us start from the first:

¹ MCPHERRAN (1987, p. 292) agrees with Burnyeat that they are cases of absolute self-refutation.

T30 Now, we have shown above that those who say that all things are false (τοὺς μὲν πάντα λέγοντας ψευδῆ) incur reversal (περιτροπομένους). For if all things are false, «All things are false», being one of all things, will be false. And if «All things are false» is false, its contradictory, «Not all things are false», will be true. Therefore, if all things are false, not all things are false. (M 8.55)

I believe that actually this passage contains strong evidence *against* interpreting περιτροπή as a form of absolute self-refutation. To begin with, it is not the proposition «All things are false» that is charged with reversal, but it is *those who say* that all things are false that are «turned about», «reversed». Given what follows in T30 and what we have learnt about περιτροπή in general, this should mean that they are reversed *into saying* that not all things are false.¹ Sextus' phrasing would be devious if all he meant were that those who maintain that all things are false are mistaken because their thesis is demonstrably false, by περιτροπή. However, one could still defend the plausible idea that such a dialectical manoeuvre is made possible by (and mirrors) a particular logical property of the thesis asserted: those who say that all things are false can be forced into admitting the contradictory of their own thesis *because* the falsehood of that thesis can be established through a proof along the lines of *Consequentia Mirabilis*.² Perhaps ancient περιτροπή always comes with an essential dialectical dress, but it would still be, intrinsically, a form of absolute self-refutation.

This conciliatory approach has its drawbacks. If we follow carefully the steps of T30's περιτροπή

- | | |
|---|--|
| (1) If all things are false, «All things are false» is false; | by universal instantiation |
| (2) If «All things are false» is false, «Not all things are false» is true; | $Fp \rightarrow T\neg p$ |
| (3) Therefore, if all things are false, not all things are false. | from (1) and (2), by transitivity and $T\neg p \rightarrow \neg p$ |

we notice that the argument halts one inference short of *Consequentia Mirabilis*: no conclusion is inferred from the conditional (3) of the form $p \rightarrow \neg p$. Should we presume that this happens because the intended conclusion, $\neg p$ («Not all things are false»), is only too clear and thus implicitly meant? On the basis of what I have argued about the shadowy status of CM in antiquity, to take such an answer for granted will not do, and we should also be cautious before interpreting (3) as a true conditional proposition which might function as the premiss of a *Consequentia Mirabilis*, given what we have learnt about TA. But, then, the proposal to catalogue this argument under the heading «absolute self-refutation» becomes questionable, and even more so when we notice that the starting move of the περιτροπή of T30 is the recognition of the self-reference of «All things are false», and not Mackie's law of T-prefixability.³

Given the *kind* of ellipticity we seem to find in Sextus, projecting the dialectical framework suggested by the first sentence of T30 into the subsequent part of the passage is not doing violence to the text. On my reading, T30 might be paraphrased as follows: «Now, as to those who assert that all things are false, we proved above that they

¹ Cf. e.g. M 8.295.

² Analogously, one might say that those who present such and such a thesis are reduced to the impossible, i.e. are bound to admit an impossibility, because their thesis does imply an impossibility.

³ For a similar argument cf. also EPICT. disc. 2.20.1-3, *pace* Barnes' prejudice that Epictetus' argument is based on CM, but is not formulated as cleanly as modern logicians would like (1997, pp. 30-31).

necessarily end up by admitting the contradictory thesis. For if *they assert that* all things are false, *they must admit that* “All things are false”, being one of all things, is itself false. And if they admit this, they cannot help conceding that its contradictory, “Not all things are false”, is true. Therefore, if they assert that all things are false, they are inexorably reversed into admitting that not all things are false.» Once again, «Everything is false» and its proponents would be unmasked as hopeless dialectical losers. But if *περιτροπή* is conceived in such a dialectical framework and its outcome is a denial of the original thesis by its own supporters, bringing CM into the picture, even supposing it possible, would be redundant.

Is such a dialectical tactic unbeatable? As I have noticed earlier, step (2) might be questionable *in this context*: although it rests on a fundamental and apparently uncontroversial principle governing *our* notions of truth and falsehood ($Fp \rightarrow T\neg p$), it is likely to be implicitly rejected by someone who dares to suggest that «Everything is false». Therefore, he could protest that that principle should not be exploited against him, question-beggingly, in a dialectical *self-refutation* argument, which should aim at defeating him by relying *only* on the consequences of his own position. How successful such a defensive manoeuvre would be is, however, far from obvious.¹

At the beginning of T30, Sextus writes as if the *περιτροπή* argument he is about to present had already been illustrated before, probably referring to this passage of the previous book:²

T31 For if all the appearances are false and nothing is true, «Nothing is true» is true. If, therefore, nothing is true, a truth exists; and in this way Xenias was driven round to the opposite of his original position (εἰς τοῦναντίον τῇ προθέσει περιήχθησαν), when he said that all appearances are false and that absolutely nothing true exists in reality. For, in general, it is impossible to say of any particular thing that is false without also affirming that something is true. For example, when we say that A is false, we are predicating the existence of that very falsity of A, and we are affirming that «A is false», so that what we virtually assert is something like this: «It is true that A is false». Simultaneously, then, with saying a thing to be false we are necessarily affirming the existence of a truth. (M 7.399)

In lieu of the usual *περιτροπή* vocabulary, we find here the phrase εἰς τοῦναντίον τῇ προθέσει περιήχθησαν. This is further valuable verification of the meaning of *περιτροπή*: «to be reversed» must be equivalent to «to be driven round to the opposite of one's original position», where one's «original position» (πρόθεσις) is what one has *said*. Once again, it is noteworthy that this turnabout is not explicitly meant to falsify Xenias' thesis, but to show why «one *must say* (λέγετον) [...] that not all <appearances> are false» (7.398).

But let us inspect T31's argument more closely. Actually its pattern differs significantly from that of T30's *περιτροπή*:³ it does not rely on self-reference,⁴ and has a form we have never encountered in ancient texts so far. Undeniably, Sextus' «If nothing is true, “Nothing is true” is true; if, therefore, nothing is true, a truth exists» bears a striking resemblance to Mackie's proof of the absolute self-refutation of «Nothing is true»:

¹ One might elicit the admission that not everything is false directly from the admission that «Everything is false» is false (by straightforward semantic descent $Fp \rightarrow \neg p$). I have explained on p. 45 that having to reject a principle like $Fp \rightarrow T\neg p$ would be a deeply embarrassing escape route; by rejecting also the principle $Fp \rightarrow \neg p$ the opponent would make his position even more unintelligible.

² So, e.g., BETT 2005, p. 99n18.

³ As BETT (2005, p. 99n18) notices. Either Sextus' backwards reference was careless, or he could not appreciate the difference between the two argument patterns.

⁴ Pace MCPHERRAN 1987, p. 292.

| <u>Mackie</u> | | <u>Sextus Empiricus</u> |
|---------------|---|---|
| (1) | $(\forall p)(p \rightarrow Tp)$ | T-prefixability |
| (2) | $\neg(\exists p)Tp \rightarrow T(\neg(\exists p)Tp)$ | From (1), by substitution |
| (3) | $T(\neg(\exists p)Tp) \rightarrow (\exists p)Tp$ | Existential generalisation |
| (4) | $\neg(\exists p)Tp \rightarrow \neg(\neg(\exists p)Tp)$ | From (2),(3), by transitivity and double negation |
| (5) | $\neg(\neg(\exists p)Tp)$ | From (4), by CM |

My contention is that this surface resemblance does not survive careful scrutiny. The problem is not only that, once again, the final inferential step by CM delivering the conclusion (5) of Mackie's absolute self-refutation does not feature in the lines of Sextus' Greek (the idea that this is fortuitous becomes less and less convincing as long as we continue stumbling across more and more cases in which the same supposed chance consistently recurs). What is even more important is that in Sextus the justification underlying (2) is *not* Mackie's T-prefixability, on the use of which in this context I have expressed my perplexity in section 2, but a different general (καθόλου) principle, according to which «it is impossible to say (λέγοντα) of any particular thing that is false without also affirming that something is true». Sextus' explanation makes it abundantly clear that this is not an abstract 'law' concerning *propositions*, but governs our ordinary practice of making *assertions* and the nature of this speech act: asserting that *p* is tantamount to asserting «“*p*” is true», i.e. making an assertion is not only predicating something of some subject but also, at the same time, positing this whole propositional content as true.¹ The structure of T₃₁'s reversal now becomes clear: when someone, like Xenias, asserts that «Nothing is true», he is, by this very speech act, unwittingly committing himself to the truth of what he is asserting; but, then, he must grant that at least one truth does exist, thus admitting the contradictory of his original position and conceding defeat. This is why one had better not *say* that «Nothing is true»; nothing is added about the truth or falsehood of the propositional content of that assertion.

Sextus' explanation of the logic of this self-refutation argument also casts new light on other passages we have analysed above. For in the case of T₃₁ it is impossible to deny that, whereas Sextus' wording of the argument could make one easily think of some kind of non-dialectical proof by CM of the logical truth of «Something is true», that formulation is actually meant to be an extremely *elliptical* reminder of the various stages of a περιτροπή which clearly presupposes a dialectical exchange in the background. But the compressed formulation of T₃₁'s περιτροπή strongly resembles, in tone and context, those we have found in T₂₀, T₂₈ and T₃₀; therefore, T₃₁ provides valuable corroboration to my proposed reading of those texts too. We started with the impression of having finally discovered an ancient instance of absolute self-refutation unequivocally matching Burnyeat's and Mackie's description; we end up with a confirmation that an-

¹ Burnyeat's claim that «the principle “There are no truths” entails its contradictory opposite» (1976a, p. 50; *italics mine*) should be emended, therefore, to «the *assertion* of the principle “There are no truths” entails its contradictory opposite», where the notion of entailment is not that of a logical consequence, but resembles that «pragmatic implication» discussed, e.g., in the literature listed in the second part of n. 2 on p. 13.

cient *περιτροπή*, in its nature and purposes, should not be identified with, or reduced to, that modern form.

It might be added that in this case the ancient approach is not only different from, but also more promising than, the modern one. For, as I have already noticed, that nothing is true is, at least on some possible understandings of the fuzzy «nothing», something which, while not *possibly-true*, is still possible.¹ In section 2 I criticised Mackie's approach for barring this logical possibility; T31's argument seems more successful in the more modest task of showing that the commonsense view that some truth exists can resist any attack in ordinary dialectical contexts,² because any denial of it will amount, at the same time, to an admission of its truth.³

7. AUGUSTINIAN TURN? THE *SOLILOQUIA* ARGUMENT FOR THE IMPERISHABILITY OF TRUTH

We are approaching the end of our journey through the ancient philosophical tradition in the hunt for early instances of self-refutation arguments against theses such as «Everything is true» and «Nothing is true». One final witness deserves our full attention, since he appears to testify to the existence of an argumentative pattern which differs from all those encountered so far and interestingly resembles that of Mackie's absolute self-refutation.

We are now at the beginning of the second book of Augustine's *Soliloquia*. In spite of its title, this short writing has the form of a dense dialogue between Augustine and a character, Ratio, who presumably represents Augustine's own reason, some kind of inner intellectual voice. The question is raised whether our souls are immortal and thus will retain (and indeed increase) their knowledge forever. As the first step towards an answer, Ratio argues for the imperishability of truth:

T31 R.: When it [*sc.* the world] will have perished, if it's going to perish, will it not then be true that the world has perished? For as long as it's not true that the world has perished, it hasn't perished. There is a conflict, therefore, between the world having perished and it not being true that the world has perished.

A.: I grant this too.

R.: And what about this: does it seem to you that it's possible that something is true, while truth does not exist?

A.: In no way.

R.: Therefore there will be truth, even if the world perishes.

A.: I cannot deny that.

R.: And what about this: if truth itself perishes, will it not be true that truth has perished? (*si ipsa veritas occidat, nonne verum erit veritatem occidisse?*)

A.: Who would deny that either?

R.: But there cannot be something true if truth doesn't exist.

¹ In such a case, the assertion that nothing is true would be Liar-paradoxical. Cf. pp. 12-13.

² A supporter of the thesis that nothing is true could object, however, that his speech acts must not be intended as carrying the ordinary import, and that the normal rules of dialectic do not hold good of them: his utterances lack assertoric force, but have some different kind of rhetorical or performative function (cf. p. 27 and p. 41n1).

³ We might say that, rather than absolutely self-refuting, «Nothing is true» is, in Mackie's own jargon, «operationally self-refuting», i.e. cannot be coherently asserted because what is implied by making an assertion contradicts the asserted content (for a similar analysis cf. PASSMORE 1961, p. 68). From this perspective, some analogy exists between the assertion of «Nothing is true» and the so-called Moorean assertions, such as «This is a table, but I do not believe that» or «This is a table, but that is not true» (cf. the literature listed in the second part of n. 2 on p. 13).

A.: I have just granted that.

R.: Therefore in no way will truth perish (*nullo modo igitur occidet veritas*). (*sol.* 2.2.2)

The elegant structure of Ratio's argument is transparent:

- (1) If truth (*veritas*) perishes, after it perishes it will be true (*verum*) that truth has perished.
- (2) If <at any time> something is true, truth will exist <at that time>.
- <3> <Even if truth perishes, truth will exist after it perishes.> From (1) and (2)
- (4) Therefore, in no way will truth perish. From <3>

Premiss (1) looks like an instantiation of a 'temporalised' version of Mackie's law of T-prefixability ($P_t x \rightarrow T_{t+1}(P_{ed} x)$): in the first half of T32 Ratio had applied the same law to «the world perishes» in order to illustrate its validity. Premiss (2) is secured by the definition of *veritas*: truth is that by which whatever is true (*verum*) is true (2.15.29), and whereas *veritas* persists when «something *verum* passes away» (1.15.28), nothing could ever be *verum* if *veritas* did not exist. It is not difficult to reconstruct how the imperishability of truth can be derived from these premisses. The unstated intermediate step <3> follows from (1) and (2), and the conclusion that *in no way* will truth perish is a straightforward consequence of it: if not even on the assumption that it perishes (i.e. ceases to exist) can *veritas* cease to exist, then there is no conceivable way *veritas* could perish. Mackie's proof of the necessary falsehood, by absolute self-refutation, of «Nothing is true» was based on T-prefixability and CM; Augustine's proof of the imperishability of truth, while not making *explicit* reference to *Consequentia Mirabilis*, seems to rely on an analogous inferential step (from <3> to (4)), and clearly starts from the application of a version of T-prefixability.¹ It seems that we have come, finally, as close as we could have hoped to a genuine antecedent for what in Mackie's analysis is an absolute self-refutation argument. Although it is presented by Ratio within her dialogue with Augustine, no dialectical context seems to be required for the argument itself to function; the vocabulary of assertion, admission, concession, so conspicuous in most of the earlier instances of περιτροπή, is absent from the various steps of Ratio's argument.² Not only would it be very hard, given the textual details, to interpret Ratio as making the point that if one asserts that *veritas* does not exist (because it has perished) then one is involuntarily and self-defeatingly conceding the existence of at least some *verum*, i.e. the content of one's own assertion, and thus the existence of *veritas* itself;³ by such a dialectical strategy she could never hope to prove what Augustine wants to be established, the impossibility that truth *will ever* perish. Ratio is trying to demonstrate a fundamental and atemporal truth about reality, and not to clarify what we can or cannot coherently think or successfully assert now. Is Augustine's innovation a real step forward? Appraisal of this issue is far from easy.⁴ To begin with, it is not obvious

¹ And *not* on self-reference, *pace* CHARRON and DOYLE, 1993, p. 247.

² That language does occur in T32, where Augustine gives his assent to the various steps of the argument proposed by Ratio, but it is not part of *the argument itself*, as the summary at 2.15.28 confirms.

³ *Contra D'* AGOSTINI 2002, pp. 45-46. Augustine employs a similar kind of dialectical strategy, instead, in his *De vera religione* (39.73): if one doubts the existence of any *verum*, and thus of *veritas*, at least one cannot doubt that one is doubting (the point that understanding, and thus presumably not doubting, that one is doubting is one of the necessary conditions for genuine doubt is made at *trin.* 10.10.14). Therefore, everyone who doubts the existence of *veritas* is certain of some *verum*, that he is doubting the existence of *veritas*. But being certain of some *verum* is inconsistent with doubt about *veritas*. For a critical analysis of this argument cf. KIRWAN 1983, pp. 219-220.

⁴ Although I shall criticise Augustine's argument, I disagree with Abercrombie's dismissive remark that «we are conscious that this is a verbal engine, of no ontological efficacy» (1938, p. 63n1).

whether we are really entitled to speak of «innovation» here: Augustine himself neither shows any awareness of his being original and reshaping an old, different dialectical manoeuvre, nor reveals any consciousness of applying a generally valid argumentative pattern (notice that he attaches no self-refutation label to the argument).

Second, when scrutinised carefully Augustine's argument betrays dangerous ambiguities, similar to those which I detected in Mackie's formulation of the absolute self-refutation of «Nothing is true» in section 2. In T₃₂ Augustine consistently uses *verum* with infinitive clauses: this might suggest, but by no means proves, that *verum* is adopted as a kind of proposition-forming operator, and not as a predicate of sentences, or whatever the truth-bearers are.¹ But how does this operator work? What is its meaning? The attempt to settle these issues on the basis of Augustine's usage of *verum* in the rest of the *Soliloquia* delivers disappointing results. At 1.15.28, shortly before T₃₂, *verum* is treated as a predicate expressing a property of *objects*: a «true tree» (*vera arbor*) is an existing, real tree, but this existential use of *verum* seems to be of little help to decide the sense of *verum* in T₃₂. This comes as no surprise when one examines the second book of the *Soliloquia* in its entirety: a very large portion of it is devoted to a dazzling quest for the meaning of *verum* and *falsum*, which ends with very few certainties. Throughout the discussion down to 2.9.16 the subjects of the predicates *verum* and *falsum* remain objects. At 2.11.19–21 disciplines, in particular grammar and dialectic, are said to be true; since disciplines include «definitions and divisions and processes of reasoning», we could welcome at this stage a first implicit recognition of the application of *verum* to propositional items. Such recognition becomes explicit only some sections later: Medea's flight on snakes, being something entirely non-existent, cannot even be called false, but the *sententia* describing it can be said to be false, and «there is a great difference between the things which are said and those about which we say something» (2.15.29). Only towards the end of the *Soliloquia* and well after T₃₂, then, are propositional items like *sententiae* (declarative sentence-tokens) explicitly accepted among the bearers of falsehood and truth: however, no account of *verum* fully relevant to its apparent use in T₃₂ has emerged.

On the basis of what we have just seen, let us examine two possible analyses of premiss (1), the kernel of Ratio's argument:

- (1a) If truth perishes, <the sentence> «Truth has perished» will then be true;
- (1b) If truth perishes, it will then be the case that truth has perished.

(1a) relies on the possibility of taking *verum* as a predicate expressing a property of *sententiae*, although we have seen that this is not the most immediate reading of Augustine's own wording. (1b) is based, instead, on a possible extension of the ontological notion of truth as real existence, applied by Augustine only to objects, to entities like facts or obtaining states of affairs: in our case, the state of affairs described by the past infinitive *veritatem occidisse* («that truth has perished»).

I suggest that on both interpretations Ratio's starting move is open to serious criticism. On reading (1a), it is not difficult to imagine a possible scenario which would falsify this conditional: on the assumption that *verum* is a predicate of sentence-tokens, if all *veritas* perished (let us say, because the whole world and mankind are annihilated),

¹ This is not to deny that «it is often difficult to determine the supposition of an expression, that is whether it is intended to be concerned with a real state of affairs, the proposition or concept formed in the mind in thinking of the state, or autonymously, with itself as a linguistic item» (CHARRON, DOYLE 1993, p. 245).

there would not be any existing sentence like «Truth has perished» thereafter, and so, *a fortiori*, no such true sentence.¹

Reading (1b) looks more promising. If *vera* are existing objects and obtaining states of affairs, the complete obliteration of reality is – one would think – what would be required for the antecedent of (1b) («Truth perishes») to become true; on such a hypothesis, however, the consequent would still be true, since it would then be the case that reality and truth have perished. Things being so, *veritas ipsa*, which, on this reading, would be the principle by which whatever is the case is the case, could never perish. Despite its apparent smoothness, this train of reasoning risks derailing, however. To begin with, the notion of ‘tensed states of affairs’ («that truth perished») to which it appeals sounds inherently odd (one could even protest that the universal annihilation presupposed by (1b)’s antecedent would not spare time, any reference to it after the hypothetical destruction of *veritas* thus becoming illicit).² A second, related concern regards a tacit presupposition on which Augustine’s argument would rely: that the past is immutable and cannot be erased. Without such a presupposition, which, although eminently reasonable and widespread in ancient thought,³ is not a logical truth, the idea that *veritas* could never become non-existent unless it first perished is no longer unassailable. For one could postulate a sudden annihilation of all the *vera*, present and past: thereafter (so to speak), it would not be the case that truth has perished, because its past existence would have been cancelled too, and what never existed cannot have perished. One might try to overcome these problems and strengthen Ratio’s argument by suggesting that «*Veritas* has perished» is not to be taken literally, as a description of a past event, but is equivalent to «*Veritas* does not exist» (or, what amounts to the same, «Nothing is *verum*»); this would make Ratio’s reasoning even more similar to Mackie’s, with the elimination of the significant difference in the use of tenses. I believe this would not be a commendable strategy, since Mackie’s own version of the argument is actually weaker than Augustine’s. If *verum* is what exists or is the case, as opposed to what does not exist or is not the case, on the hypothesis that *veritas* does not exist and nothing is *verum* reality is, so to speak, a complete blank. To postulate the existence of some kind of ‘second-order *verum*’, that nothing (or, avoiding self-reference, nothing else) is *verum*, means to empty *verum* and *veritas* of their supposed role: although of course nothing precludes this kind of prefixability *from a merely formal point of view*, we must consider what it would amount to. Consider the following example. Socrates is not alive, does no longer exist (Augustine would perhaps be ready to say that Socrates is not *verus*); isn’t there a state of affairs which now obtains, i.e. Socrates not being alive? One might reasonably reply that there is no state of affairs which obtains today which makes the sentence «Socrates is alive» false: «Socrates is alive» is false today because the state of affairs that Socrates is alive does not obtain, and not because some mysterious negative state of affairs (that Socrates is *not* alive) obtains. To attach the adjective *verum* to negative states of affairs, albeit formally unimpeachable, is to empty it of the meaning and function we are attributing to it here: for, as a consequence of this move, *veritas* would become a principle of both existence (it is because of *veritas* that it is the case that I am

¹ WATSON (1990, p. 184) takes reading (1a) for granted and proposes an analogous criticism. This objection is similar to that which I have formulated in sect. 2 against Mackie’s argument on reading (b).

² Augustine could have been sympathetic towards such a line of reasoning, if we judge by his argument for the non-eternity of time in *Confessions* 11 and *De civitate* 11, according to which it makes no sense to ask what God did *before* creating the world and time (cf. p. 67n4).

³ Cf. e.g. ARIST. *NE* VI 2, 1139b8–11.

alive) and of non-existence (it is because of *veritas* that it is the case that Socrates is not alive). Moreover, to crowd one's ontology with a swarm of negative states of affairs obtaining (being *vera*) whenever the corresponding positive ones do not obtain seems to be an unavoidable toll for Augustine's argument on this reading of its main premiss:¹ but in the presence of such an unparsimonious ontology we need not be fanatic Ockhamists to shiver (not to speak of the oddness of describing non-linguistic items like state of affairs as 'negative').² Ratio's move, on interpretation (1b) of its crucial premiss, would be questionable also for a contextual reason. In the *Soliloquia* Augustine's ultimate goal when arguing for the imperishability of *veritas* is establishing the immortality of the soul: if *veritas* is imperishable, and *veritas* must dwell, inseparably, in our souls, then our souls must themselves be immortal (2.19.33). But the existence of that *veritas* established on interpretation (1b), i.e. the existence of at least some *verum* in the formal sense just discussed, cannot guarantee the existence of souls, since that only *verum* which is guaranteed to exist forever is an entity which need not (and perhaps cannot) dwell in our souls (the supposedly true fact that everything, including us, has been annihilated). Augustine seems to offer some revision, when he claims, late in the second book, that bona fide *vera* are only the objects of the *disciplinae*, e.g. geometrical entities like squares and circles. But this idea is impossible to reconcile with the use of *verum* which Ratio has made in her self-refutation argument in T32. If that argument works, it guarantees the eternal existence of one and only one *verum* which is irrelevant to the proof of the immortality of the soul: for its validity would be perfectly consistent with a scenario in which the whole of reality has been obliterated, and with it all the *vera*, except for one single *verum*, the fact that every other *verum* has perished. To make a comparison, the inscription on a sheet of paper «All writing has been cancelled from this sheet» cannot be true, being itself a surviving instance of writing, but this by no means proves that there must be some *other* inscription on that sheet of paper (or that it is impossible to erase all the other inscriptions originally written on it).³

Augustine himself might have been aware of at least some of the difficulties of Ratio's strategy I have denounced. After a summary of T32 at 2.15.28, his plea for more time to assess the merits of the argument and his promise that, although he himself could not find any objection, he will make sure «that learned and prudent men read these things and correct any rashness of ours there may be» could be interpreted as hints at some genuine perplexity or dissatisfaction.⁴

¹ Notice that Mackie faces analogous problems on some interpretations of his T-prefixability law (cf. sect. 2).

² For example, on the basis of *Metaph.* Δ 29, 1024b17-21 Crivelli convincingly argues that «the only states of affairs recognised by Aristotle are "affirmative" states of affairs» (2004b, pp. 49-50). This is not to say that an ontology which allows the existence of negative facts (obtaining negative states of affairs) corresponding to false propositions and making them false is sheer nonsense. Such an ontology was defended, for example, by Russell in his 1918 lectures on logical atomism (cf. RUSSELL 1956, p. 211).

³ The parallel is not perfect, because in my example also the sentence «All writing has been cancelled from this sheet» could be cancelled, whereas the pivot of our self-refutation argument is precisely that the fact that *veritas* has perished could never be cancelled even if *veritas* perished.

⁴ Ratio's argument exhibits some structural resemblance with this argument for the eternity of time: if time was created, there was a time before then in which no time existed; if time will perish, there will be a time after then in which time will not exist; therefore time neither was created nor will perish (cf. e.g. S.E. M 10.189). Since Augustine rejected this line of reasoning (cf. e.g. *civ. dei* 11.4-6; 12.16; *conf.* 11.13.15; 11.30.40), he could have been doubtful about endorsing something analogous about truth.

Augustine's appeal to *docti atque prudentes viri* could not lapse unheard: in the Middle Ages a handful of illustrious readers of his *Soliloquia* eagerly returned to Ratio's self-refutation argument, either to endorse it (e.g.

8. CONCLUSION

We have finally reached the end of our exploration of the ancient history and logic of the self-refutation charge against «Everything is true», «Everything is false», and cognate theses.¹ We started from Mackie's influential account of absolute self-refutation, on the form and content of which I have expressed some perplexities (sect. 2). We have then discovered that, contrary to what is often maintained in the literature, Mackie's approach based on T-prefixability and *Consequentia Mirabilis* finds no clear parallel in the ancient texts, with the unique, remarkable and late exception of Augustine's *Soliloquia* (sect. 7).² I have argued that this conspicuous absence does not betray a defect of logical rigour which prevented the ancients from achieving full consciousness of the logical form of their self-refutation arguments, and thus from moulding them into more precise, fully formalised Mackie-style shape. Through careful re-examination of the textual evidence in sects. 3-6 I have tried to prove something which has been obscured by the uncritical adoption of the modern paradigm as the only guidance to our interpretation of the ancient testimonies: the ancient approach to self-refutation, which only from the Hellenistic age came to be widely identified by the label *περιτροπή* but has been revealed to be quite unitary in some of its basic features, is not a muddled ancestor of the modern one, but differs from it in philosophically interesting ways. Not only was Mackie's strategy in the pattern of *Consequentia Mirabilis* never adopted before Augustine as a matter of fact, but it *could not* have been adopted by anyone accepting certain fundamental 'non-classical' features entrenched in the two most prominent ancient logical systems, the Aristotelian and the Chrysippean (sect. 6). Ancient self-refutation arguments do not aim at establishing, *in vacuo*, the truth-value (necessary falsehood) of the propositional content of theses like «Everything is true» or «Nothing is true»; they criticise such theses as dialectical losers, because whoever dares to propose and defend them can be forced into admitting their contradictories and rejecting them as a direct consequence of proposing or trying to defend them. In other terms, the ancient self-refutation charge, at least in so far as it was applied to the kind of 'extremist' theses which have been the subject-matter of our inquiry here,³ does not aim at establishing by logic some absolute truth about the world, but at clarifying what can and can-

Anselm, Thomas) or to criticise it (e.g. Scotus). I cannot analyse this fascinating history here (for which cf. CHARRON, DOYLE 1993 and CASTAGNOLI 2005). I only signal that

(1) one of the keenest supporters of the argument, Bonaventure of Bagnoregio, seems to misinterpret it in such a way as to make it extraordinarily similar to Sextus' *different* dialectical argument in T31 (cf. e.g. *Myst. Trinit.* Q. 1, art. 1, ad n. 5 *e contrario*; contra CHARRON, DOYLE 1993, pp. 250-252);

(2) at *Ordin.* 1, d. 2, p. 1, q. 2, n. 45 John Duns Scotus rejected a proof for the necessary existence of *veritas* similar to Augustine's with considerations analogous to those I have illustrated above. The idea that if nothing existed and nothing were true then there would be no intellect which could truly judge this (and thus no bearer of what Scotus calls «formal truth») matches my criticism of the soundness of Ratio's argument on the assumption that *verum* is a predicate of sentence-tokens. Scotus' remark that if nothing existed then *veritas in re* (identified with existence and «founding» reality) would not exist either seems to be based on a tacit refusal to attribute any positive ontological status, deserving the label *verum*, to wholly negative facts like the fact that nothing exists.

¹ For an overview of the same subject in medieval and modern times cf. D'AGOSTINI 2002.

² This exception is even more significant in the light of the fact that Augustine fathered another non-dialectical self-refutation argument, the famous «*Si fallor, sum*», based on an implicit application of CM (cf. *lib. arb.* 2.3.7; *civ. dei* 11.26; *trin.* 15.12.21), on which cf. MATTHEWS 1992, pp. 32-33, HARRISON 1999 and CASTAGNOLI 2005.

³ I only mention, dogmatically, that my broader research in CASTAGNOLI 2005 showed that the same conclusion holds good for ancient self-refutation arguments quite generally.

not be successfully entertained in dialectical exchange, the original *locus* of philosophical inquiry. The dialectical context, thus, is not simply the broad and natural *background* on which self-refutation arguments happened to be formulated, but it is typically a *necessary condition* for them to work properly. Although ancient self-refutation arguments cannot 'falsify' our most radical adversaries' views (and defuse our own most hyperbolic doubts) by proving that what they envisage is 'logically impossible', they can silence them, by delimiting the area of constructive philosophical inquiry and debate. And even if silencing your adversary does not necessarily amount to proving the truth of your own position, it can be something extremely valuable if your position is already the default one, and therefore you do not need to win new ground, but only to withstand the siege. However, the self-refutation arguments themselves were not perceived as philosophical wonders by the ancients; it is the self-refuting positions that were seen as amazing in their hopelessness.

A thorny question is what level of consciousness we are entitled to attribute to the *ancients themselves* of anything like my distinction between an absolute proof of the falsehood of a certain proposition and a dialectical silencer of its supporters. Although no such distinction is explicitly articulated in any of our sources, I have suggested that it might be reflected in the caution with which the results of self-refutation are often cast in terms of the admissions or concessions to which the proponents of certain theses are finally bound. However, we have also noticed an opposite tendency to elliptical formulations which are likely to blur that distinction: it remains hard to decide whether such a tendency reflects some confusion between the two distinct levels or testifies exactly to the opposite, i.e. a dialectical setting for self-refutation was so obvious that making it fully explicit was sometimes felt as unnecessary.

I have praised the ancient approach to self-refutation for its 'modesty' in setting itself a dialectical goal which can be achieved with available dialectical resources. This does not mean that all the self-refutation manoeuvres we have encountered in sects. 3-6 easily fulfil their proposed task. A supposed advantage of self-refutation arguments is that they exploit only what one's opponent has *already* admitted, or is *already* committed to conceding, in virtue of his presenting his own position; to agree on the precise extent of one's dialectical commitments, however, can be no less difficult than agreeing on the truth-value of the premisses of an ordinary direct refutation. I have often signalled the possibility for the target of the self-refutation charge to protest that the argument deployed against him is actually question-beggingly foisting upon him tacit assumptions to which not only has he never committed himself, either explicitly or implicitly, but which he has actually rejected, at least by implication, in the very act of presenting his revisionist views. The idea that all self-refutation arguments could, or should, avail themselves *exclusively* of the self-refuting thesis is illusory; their success will depend on the degree to which the further presuppositions involved are perceived by all the parties to the debate (including the audience) as the immovable background conditions for any discussion on a certain subject to take place and remain an intelligible and genuine discussion on *that* subject, and, correspondingly, on the degree to which any attack on these presuppositions is bound to appear as a desperate and merely *ad hoc* tactic to avoid admitting an only too clear defeat.

REFERENCES

- ABERCROMBIE N., *Saint Augustine and French Classical Thought*, Oxford, Clarendon Press, 1938.
- AUBENQUE P., *Sur l'inauthenticité du livre K de la Métaphysique*, in P. Moraux, J. Wiesner (eds.), *Zweifelhaftes im Corpus Aristotelicum: Studien zu einigen Dubia*. Akten des 9. Symposium Aristotelicum (Berlin, 7-16 September 1981), Berlin-New York, de Gruyter, 1983, pp. 318-344.
- BALTZLY D. C., *Aristotle and Platonic Dialectic in Metaphysics I'4*, «Apeiron», 32 (1999), pp. 171-202.
- BARNES J., *The catalogue of Chrysippus' logical works*, in K. A. Algra, P. W. van der Horst, D. T. Runia (eds.), *Polyhistor: Studies in the History and Historiography of Ancient Philosophy*, Leiden-New York-Köln, Brill, 1996, pp. 169-184.
- BARNES J., *Logic and the Imperial Stoa*, Leiden-New York-Köln, Brill, 1997.
- BARNES J., *Argument in Ancient Philosophy*, in D. Sedley (ed.), *The Cambridge Companion to Greek and Roman Philosophy*, Cambridge, Cambridge University Press, 2003, pp. 20-41.
- BARNES J., *What is a disjunction?*, in D. Frede, B. Inwood (eds.), *Language and Learning: Philosophy of Language in the Hellenistic Age*, Cambridge, Cambridge University Press, 2005, pp. 274-298.
- BELLISSIMA F., PAGLI P., *Consequentia Mirabilis. Una regola logica tra matematica e filosofia*, Firenze, Olschki, 1996.
- BEMELMANS R., *Why Does Protagoras Rush Off? Self-Refutation and Haste in Plato, Theaetetus 169a-171d*, «Ancient Philosophy», 22 (2002), pp. 75-86.
- BETT R., *Sextus Empiricus. Against the Logicians*, Cambridge, Cambridge University Press, 2005.
- BOBZIEN S., *Stoic Syllogistic*, «Oxford Studies in Ancient Philosophy», 14 (1996), pp. 133-192.
- BOLTON R., *Metaphysics as a Science*, in T. Scaltsas, D. Charles, M. L. Gill (eds.), *Unity, Identity, and Explanation in Aristotle's Metaphysics*, Oxford, Clarendon Press, 1994, pp. 321-354.
- BONNEY W. L., *Operational Self-Refutation*, «Philosophical Quarterly», 16 (1966), pp. 348-351.
- BOSTOCK D., *Plato's Theaetetus*, Oxford, Clarendon Press, 1988.
- BOYLE J. M. Jr., *Self-Referential Inconsistency, Inevitable Falsity and Metaphysical Argumentation*, «Metaphilosophy», 3 (1972), pp. 25-42.
- BURNET J., *Platonis Opera: Tomus I, tetralogias I-II continens*, Oxford, Oxford University Press, 1900.
- BURNYEAT M. F., *Protagoras and Self-Refutation in Later Greek Philosophy*, «Philosophical Review», 85 (1976), pp. 44-69. [1976a]
- BURNYEAT M. F., *Protagoras and Self-Refutation in Plato's Theaetetus*, «Philosophical Review», 85 (1976), pp. 172-195. [1976b]
- BURNYEAT M., *The Theaetetus of Plato*, translation by M. J. Levett, Indianapolis-Cambridge, Hackett, 1990.
- BURNYEAT M. F., *Antipater and Self-Refutation: Elusive Arguments in Cicero's Academica*, in B. Inwood, J. Mansfeld (eds.), *Assent and Argument: Studies in Cicero's Academic Books*. Proceedings of the 7th Symposium Hellenisticum (Utrecht, August 21-25, 1995), Leiden, Brill, 1997, pp. 277-310.
- BURNYEAT M. F., *Dissoi logoi*, in E. Craig (ed.), *Routledge Encyclopaedia of Philosophy in CD-ROM*, London, Routledge, 1998.
- BURNYEAT M. F., *Plato on how not to speak of what is not: Euthydemus 283a-288a*, in P. Pellegrin, M. Canto-Sperber (eds.), *Le Style de la Pensée*, Paris, Belles Lettres, 2002, pp. 40-66.
- CAMPBELL L., *The Theaetetus of Plato*, Oxford, Clarendon Press, 1883².
- CASSIN B., *Il senso di Gamma e la strategia di Aristotele contro i Presocratici in Metafisica IV*, «Rivista di Filosofia Neo-scolastica», 85 (1993), pp. 533-565.
- CASSIN B., NARCY M., *La décision du sense: Le livre Gamma de la Métaphysique d'Aristote*, Paris, Vrin, 1989.
- CASTAGNOLI L., *Self-bracketing Pyrrhonism*, «Oxford Studies in Ancient Philosophy», 18 (2000), pp. 263-328.
- CASTAGNOLI L., *Protagoras Refuted: How Clever is Socrates' "Most Clever" Argument at Theaetetus 171a-c?*, «Topoi», 23 (2004), pp. 3-32. [2004a]
- CASTAGNOLI L., *Review of Nasti 2002*, «Elenchos», 25 (2004), pp. 182-195. [2004b]

- CASTAGNOLI L., *Il condizionale crisippeo e le sue interpretazioni moderne*, «Elenchos», 25 (2004), pp. 353-395. [2004c]
- CASTAGNOLI L., *The Logic of Ancient Self-Refutation: From Democritus to Augustine*, Ph.D. thesis, Cambridge, 2005 (University of Cambridge Library and Faculty of Classics Library).
- CASTAGNOLI L., *Συνάφτης crisippea e tesi di Aristotele*, in *La logica nel pensiero antico* (Atti del I Colloquio, Roma, 28-29 Novembre 2000), Napoli, Bibliopolis, 2008 (forthcoming).
- CAVINI W., *Chrysippus on Speaking Truly and the Liar*, in K. Döring, T. Ebert (eds.), *Dialektiker und Stoiker: Zur Logik der Stoa and ihrer Vorläufer*, Stuttgart, Steiner, 1993, pp. 85-109.
- CAVINI W., *Principia Contradictionis: Sui Principi Aristotelici della Contraddizione (§§ 1-3)*, «Antiquorum Philosophia», 1 (2007), pp. 123-169.
- CAVINI W., *Principia Contradictionis: Sui Principi Aristotelici della Contraddizione (§§ 4)*, «Antiquorum Philosophia», 2 (2008) (forthcoming).
- CHAMPLIN T. S., *Reflexive Paradoxes*, London-New York, Routledge, 1988.
- CHANCE T. H., *Plato's Euthydemus: Analysis of What Is and Is not Philosophy*, Berkeley-Los Angeles-Oxford, University of California Press, 1992.
- CHAPPELL T. D. J., *Reading Plato's Theaetetus*, Indianapolis-Cambridge, Hackett, 2005.
- CHAPPELL T. D. J., *Reading the περίσοφίη: Theaetetus 170c-171c*, «Phronesis», 51 (2006), pp. 109-139.
- CHARRON W., DOYLE J. P., *On the Self-Refuting Statement "There is No Truth": A Medieval Treatment*, «Vivarium», 31 (1993), pp. 241-266.
- CHERNISS H. F., *Aristotle's Criticism of Presocratic Philosophy*, Baltimore, John's Hopkins, 1935.
- CONLEY T. M., *Dating the so-called Dissoi Logoi: A Cautionary Note*, «Ancient Philosophy», 5 (1985), pp. 59-65.
- COOPER G. L. III, *Attic Greek Prose Syntax*, after K. W. Krüger, 2 vols., Ann Arbor, University of Michigan Press, 1988.
- CORNFORD F. M., *Plato's Theory of Knowledge: The Theaetetus and the Sophist of Plato translated with a running commentary*, London, Routledge & Kegan Paul, 1935.
- CRIVELLI P., *Aristotle on the Liar*, «Topoi», 23 (2004), pp. 61-70. [2004a]
- CRIVELLI P., *Aristotle on Truth*, Cambridge, Cambridge University Press, 2004. [2004b]
- D'AGOSTINI F., *Disavventure della verità*, Torino, Einaudi, 2002.
- DANCY R. M., *Sense and Contradiction: A Study in Aristotle*, Dordrecht, Reidel, 1975.
- DE PRAETERE T., *The Demonstration by Refutation of the Principle of Non-Contradiction in Aristotle's Metaphysics, Book IV*, «Logique et Analyse», 36 (1993), pp. 343-358.
- DENYER N., *Language, Thought and Falsehood in Ancient Greek Philosophy*, London-New York, Routledge, 1991.
- DIÈS A., *Platon: Oeuvres complète, Tome VIII-2^e Partie: Théétète*, Paris, Les Belles Lettres, 1924.
- DUKE E. A., HICKEN W. F., NICOLL W. S. M., ROBINSON D. B., STRACHAN J. C. G., *Platonis Opera: Tomus I, tetralogias I-II continens*, Oxford, Oxford University Press, 1995.
- EMILSSON E. K., *Plato's Self-Refutation Argument in Theaetetus 171A-C Revisited*, «Phronesis», 39 (1994), pp. 136-149.
- FINE G., *Plato's Refutation of Protagoras in the Theaetetus*, «Apeiron», 31 (1998), pp. 201-234.
- FINNIS J. M., *Scepticism, Self-Refutation, and the Good of Truth*, in P. M. S. Hacker, J. Raz (eds.), *Law, Morality, and Society*, Oxford, Clarendon Press, 1977, pp. 247-267.
- FORD A., *Protagoras' Head: Interpreting Philosophic Fragments in the Theaetetus*, «American Journal of Philology», 115 (1994), pp. 199-218.
- GOTTLIEB P., *The Principle of Non-Contradiction and Protagoras: The Strategy of Aristotle's Metaphysics IV 4*, «Proceedings of the Boston Area Colloquium in Ancient Philosophy», 8 (1992), pp. 183-198.
- GRAHAM A. C., *Disputers of the Tao: Philosophical Arguments in Ancient China*, LaSalle, Open Court, 1989.
- GRANT C. K., *Pragmatic Implication*, «Philosophy», 33 (1958), pp. 303-324.
- GREEN M. S., WILLIAMS J. N. (eds.), *Moore's Paradox: New Essays on Belief, Rationality, and the First Person*, Oxford, Oxford University Press, 2007.

- HARBSMEIER C., *Science and Civilisation in China, Vol. 7, Part 1: Language and Logic*, Cambridge, Cambridge University Press, 1998.
- HARRISON S., *Do we have a Will? Augustine's Way in to the Will*, in G. B. Matthews (ed.), *The Augustinian Tradition*, Berkeley-Los Angeles-London, University of California Press, 1999, pp. 195-205.
- HAWTREY R. S. W., *Commentary on Plato's Euthydemus*, Ann Arbor, Edwards Brothers, 1981.
- HERRNSTEIN SMITH B., *Unloading the Self-Refutation Charge*, in R. T. Ames, W. Dissanayake (eds.), *Self and Deception: A Cross-Cultural Philosophical Enquiry*, Albany, SUNY Press, 1996, pp. 143-160.
- HINTIKKA J., *Knowledge and Belief*, Ithaca, Cornell University Press, 1962.
- HUGHES G. E., *John Buridan on Self-reference: Chapter Eight of Buridan's Sophismata, with a Translation, an Introduction, and a philosophical Commentary*, Cambridge, Cambridge University Press, 1982.
- IRWIN T. H., *Aristotle's First Principles*, Oxford, Clarendon Press, 1988.
- JAEGER W., *Aristotle: Fundamentals of the History of his Development*, Oxford-London, Oxford University Press, 1934.
- JOHANSSON I., *Performatives and Antiperformatives*, «Linguistics and Philosophy», 26 (2003), pp. 661-702.
- JOHNSTONE H. W. jr., *Self-Refutation and Validity*, «Monist», 48 (1964), pp. 467-485.
- JOHNSTONE H. W. jr., *Self-Application in Philosophical Argumentation*, «Metaphilosophy», 20 (1989), pp. 247-261.
- JØRGENSEN J., *Some Reflections on Reflexivity*, «Mind», 63 (1953), pp. 289-300.
- KAHN C. H., *Some Puzzles in the Euthydemus*, in T. M. Robinson, L. Brisson (eds.), *Plato: Euthydemus, Lysis, Charmides. Proceedings of the v Symposium Platonicum*, Sankt Augustin, Academia Verlag, 2000, pp. 88-97.
- KETCHUM R. J., *Plato's 'Refutation' of Protagorean Relativism: Theaetetus 170-171*, «Oxford Studies in Ancient Philosophy», 10 (1992), pp. 73-105.
- KIRWAN C., *Augustine against the Sceptics*, in M. Burnyeat (ed.), *The Skeptical Tradition*, Berkeley-Los Angeles, University of California Press, 1983, pp. 205-223.
- KNEALE W., *Aristotle and the Consequentia Mirabilis*, «Journal of Hellenic Studies», 77 (1957), pp. 62-66.
- KOYRÉ A., *The Liar*, «Philosophy and Phenomenological Research», 6 (1946), pp. 344-362.
- LEAR J., *Aristotle and Logical Theory*, Cambridge, Cambridge University Press, 1980.
- LEE E. N., *Hoist With his Own Petard: Ironical and Comic Elements in Plato's Critique of Protagoras (Tht. 161-171)*, in E. N. Lee, A. P. D. Mourelatos, R. M. Rorty (eds.), *Exegesis and Argument: Studies in Greek Philosophy presented to Gregory Vlastos*, Assen, Van Gorcum, 1973, pp. 225-261.
- LEE M.-K., *Epistemology After Protagoras: Responses to Relativism in Plato, Aristotle, and Democritus*, Oxford, Clarendon Press, 2005.
- LEVI A. J., *On Twofold Statements*, «American Journal of Philology», 61 (1940), pp. 292-306.
- LONG A., *Refutation and Relativism in Theaetetus 161-171*, «Phronesis», 49 (2004), pp. 24-40.
- ŁUKASIEWICZ J., *Aristotle's Syllogistic: From the Standpoint of Modern Formal Logic*, Oxford, Clarendon Press, 1957².
- ŁUKASIEWICZ J., *Remarks on Many-Valued Systems of Propositional Logic*, in O. Wojtasiewicz (ed.), *Jan Łukasiewicz. Selected Works*, Amsterdam-London, North-Holland, 1970, 153-178 [English translation of *Philosophische Bemerkungen zu den mehrwertigen Systemen des Aussagenkalkulus*, «Monatshefte für Mathematik und Physik», 37 (1930), pp. 51-77].
- MACKIE J. L., *Self-Refutation – A Formal Analysis*, «Philosophical Quarterly», 14 (1964), pp. 193-203 [reprinted in J. L. Mackie, P. Mackie (eds.), *Logic and Knowledge*, Oxford, Clarendon Press, 1985, pp. 54-67].
- MATTHEWS G. B., *Thought's Ego in Augustine and Descartes*, Ithaca-London, Cornell University Press, 1992.
- MCCABE M. M., *Plato's Individuals*, Princeton, Princeton University Press, 1994.
- MCCABE M. M., *Silencing the Sophists: The Drama of the Euthydemus*, «Proceedings of the Boston Area Colloquium in Ancient Philosophy», 14 (1998), pp. 139-168.

- MCCABE M. M., *Plato and His Predecessors: The Dramatisation of Reason*, Cambridge, Cambridge University Press, 2000.
- MCCALL S., *Connexive Implication*, «Journal of Symbolic Logic», 31 (1966), pp. 415-433.
- MCDOWELL J., *Plato. Theaetetus*, Oxford, Clarendon Press, 1973.
- MCPHERRAN M. L., *Skeptical Homeopathy and Self-refutation*, «Phronesis», 32 (1987), pp. 290-328.
- MIGNUCCI M., *Aristotele. Analitici Primi*, Napoli, Loffredo, 1969.
- MIGNUCCI M., *The Liar Paradox and the Stoics*, in K. Ierodiakonou (ed.), *Topics in Stoic Philosophy*, Oxford, Clarendon Press, 1999, pp. 54-70.
- MOORE G. E., *Reply to my Critics*, in P. A. Schilpp (ed.), *The Philosophy of G. E. Moore*, Evanston (Ill.), Northwestern University, 1942.
- MOREL P.-M., *Démocrite et la recherche des causes*, Paris, Klincksieck, 1996.
- NARCY M., *A qui la parole? Platon et Aristote face à Protagoras*, in B. Cassin (ed.), *Positions de la sophistique*, Paris, Vrin, 1986, pp. 75-90.
- NARCY M., *Platon revu et corrigé*, in B. CASSIN, M. NARCY, *La décision du sens: Le livre Gamma de la Métaphysique d'Aristote*, Paris, Vrin, 1989, pp. 61-110.
- NARCY M., *Platon: Théétète*, Paris, Flammarion, 1995².
- NASTI DE VINCENTIS M., *Logica Scettica e Implicazione Stoica*, in G. Giannantoni (ed.), *Lo scetticismo antico*. Atti del convegno organizzato dal «Centro di Studio del Pensiero Antico» del C.N.R. (Roma, 5-8 novembre 1980), 2 vols., Napoli, Bibliopolis, 1981, vol. II, pp. 501-532.
- NASTI DE VINCENTIS M., *Logiche della connessività*, Bern, Haupt, 2002.
- NEWMAN J., *The Recoil Argument*, «Apeiron», 16 (1982), pp. 47-52.
- NUCHELMANS G., *Theories of the proposition: Ancient and medieval conceptions of the bearers of truth and falsity*, Amsterdam-London, North Holland, 1973.
- NUCHELMANS G., *Dilemmatic arguments: Towards a history of their logic and rhetoric*, Amsterdam-Oxford, North Holland, 1991.
- PAGE C., *On Being False by Self-Refutation*, «Metaphilosophy», 23 (1992), pp. 410-426.
- PASSMORE J. A., *Philosophical Reasoning*, New York, Scribner's, 1961.
- PATZIG G., *Aristotle and Syllogisms from False Premisses*, «Mind», 68 (1959), pp. 186-192.
- POLANSKY R. M., *Philosophy and Knowledge: A Commentary on Plato's Theaetetus*, Lewisburg-London, Bucknell University Press-Associated University Press, 1992.
- POLITIS V., *Aristotle and the Metaphysics*, London-New York, Routledge, 2004.
- PRIOR A. N., *Epimenides the Cretan*, «Journal of Symbolic Logic», 23 (1958), pp. 261-266.
- PRIOR A. N., *The Possibly-True and the Possible*, «Mind», 78 (1969), pp. 481-492.
- RANKIN H. D., 'Οὐκ ἔστιν ἀντιλέγειν', in G. B. Kerferd (ed.), *The Sophists and Their Legacy*, Wiesbaden, Steiner, 1981, pp. 25-37.
- REALE G., *Aristotele. La Metafisica*, 3 vols., Milano, Vita e Pensiero, 1993.
- ROBINSON T. M., *Contrasting Arguments: An Edition of the Dissoi Logoi*, New York, Arno Press, 1979.
- ROSS W. D., *Aristotle's Metaphysics*, 2 vols., Oxford, Clarendon Press, 1924.
- RUSSELL B., *The Philosophy of Logical Atomism*, in R. C. Marsh (ed.), *Logic and Knowledge: Essays 1902-1950*, London, Allen & Unwin, 1956, pp. 177-281.
- RÜSTOW A., *Der Lügner: Theorie, Geschichte, und Auflösung*, Leipzig, Teubner, 1910.
- SCHREIBER S. G., *Aristotle on False Reasoning: Language and World in the Sophistical Refutations*, Albany, SUNY Press, 2003.
- SEDLEY D. N., *The Midwife of Platonism: Text and Subtext in Plato's Theaetetus*, Oxford, Clarendon Press, 2004.
- SMITH R., *Aristotle's Prior Analytics*, Indianapolis, Hackett, 1989.
- SORENSEN R. A., *Blindspots*, Oxford, Clarendon Press, 1988.
- SPRAGUE R. K., *Plato's use of fallacy: A study of the Euthydemus and some other dialogues*, London, Routledge & Kegan Paul, 1962.
- STACK M., *Self-Refuting Arguments*, «Metaphilosophy», 14 (1983), pp. 327-335.
- STOPPER M. R., *Schizzi Pirroniani*, «Phronesis», 28 (1983), pp. 265-297.

- STROUD B., *Transcendental Arguments*, «Journal of the History of Philosophy», 65 (1968), pp. 241-256.
- VAILATI G., *A proposito di un passo del Teeteto e di una dimostrazione di Euclide*, «Rivista di filosofia e scienze affini», 6 (1904), pp. 5-15.
- VANDERVEKEN D., *Illocutionary Logic and Self-Defeating Speech Acts*, in J. R. Searle, F. Kiefer, M. Bierwisch (eds.), *Speech Act Theory and Pragmatics*, Dordrecht, Reidel, 1980, pp. 247-272.
- WATERFIELD R. A. H., *Plato. Theaetetus*, Harmondsworth, Penguin, 1973.
- WATERLOW S., *Protagoras and Inconsistency: Theaetetus 171a6-c7*, «Archiv für Geschichte der Philosophie», 59 (1977), pp. 19-36.
- WATSON G., *Saint Augustine. Soliloquies and Immortality of the Soul*, Warminster, Aris & Phillips, 1990.
- WEDIN M. V., *Some Logical Problems in Metaphysics I*, «Oxford Studies in Ancient Philosophy», 19 (2000), pp. 113-162.
- WEDIN M. V., *A Curious Turn in Metaphysics Gamma: Protagoras and Strong Denial of the Principle of Non-Contradiction*, «Archiv für Geschichte der Philosophie», 85 (2003), pp. 107-130.
- WEDIN M. V., *Animadversions on Burnyeat's Theaetetus: On the Logic of the Exquisite Argument*, «Oxford Studies in Ancient Philosophy», 29 (2005), pp. 171-191.
- WHITE F. C., *Self-refuting Propositions and Relativism*, «Metaphilosophy», 20 (1989), pp. 84-92.
- WOHLRAB M., *Platonis Theaetetus*, Leipzig, Teubner, 1891.